5-5-2016

Evolution and Biological Roles of Three-Finger Toxins in Snake Venoms

Cassandra M. Modahl

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

EVOLUTION AND BIOLOGICAL ROLES OF THREE-FINGER TOXINS IN SNAKE VENOMS

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Cassandra M. Modahl

College of Natural and Health Sciences
School of Biological Sciences
Biological Education

May 2016

This Dissertation by: Cassandra M. Modahl

Entitled: *Evolution and Biological Roles of Three-finger Toxins in Snake Venoms*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Natural and Health Sciences in School of Biological Sciences, Program of Biological Education

Accepted by the Doctoral Committee

_____
Stephen Mackessy Ph.D., Research Advisor

_____
Susan Keenan Ph.D., Committee Member

_____
Mitchell McGlaughlin Ph.D., Committee Member

_____
Seth Frietze Ph.D., Committee Member

_____
Aichun Dong Ph.D., Faculty Representative

Date of Dissertation Defense _____

Accepted by the Graduate School

_____
Linda L. Black, Ed.D.
Associate Provost, Dean of the Graduate School and International Admissions

ABSTRACT

Modahl, Cassandra M. *Evolution and Biological Roles of Three-finger Toxins in Snake Venoms.* Published Doctor of Philosophy dissertation, University of Northern Colorado, 2016.

Snake venoms are complex mixtures of many enzymatic and non-enzymatic proteins, as well as small peptides. Several major venom protein superfamilies, including three-finger toxins, phospholipases A_2, serine proteinases, metalloproteinases, proteinase inhibitors and lectins, are found in almost all snake venoms, from front-fanged viperids (vipers and pit vipers) and elapids (cobras, mambas, sea snakes, etc.) to rear-fanged colubrids. However, these proteins vary in abundance and functionality between species.

Variation in snake venom composition is attributed to both differences in the expression levels of toxin encoding genes and occurrence of amino acid sequence polymorphisms. Documenting intraspecific venom variation has both clinical (antiserum development) and biological (predator and prey coevolution) implications. Venom is primarily a trophic adaptation and as such, the evolution and abundance of venom proteins relates directly to prey capture success and organism natural history. Without this biologically relevant perspective, proteomic and transcriptomic approaches could produce simply a list of proteins, peptides, and transcripts. It is therefore important to consider the presence and evolution of venom proteins in terms of their biological significance to the organism.

Three-finger toxins (3FTx) comprise a particularly common venom protein superfamily that contributes significantly to differences in envenomation symptomology, toxicity, and overall venom composition. Three-finger toxins are non-enzymatic proteins that maintain a common molecular scaffold, and bind to different receptors/acceptors and exhibit a wide variety of biological effects. These toxins are the main lethal neurotoxins in some snake venoms and are currently the only known venom proteins associated with prey-specific toxicity.

This dissertation has four major objectives: (i) to examine 3FTxs in front-fanged Elapidae and rear-fanged snake venoms for prey-specific toxicity, (ii) to examine differences in 3FTx expression within rear-fanged snake venom glands, (iii) to determine if mRNA transcripts obtained from crude venoms can be utilized for molecular evolutionary studies and venom proteomic studies, and (iv) to determine if a transcriptomic and proteomic integrated approach can more thoroughly characterize differences in rear-fanged snake venom composition.

Three-finger toxins were isolated from the venom of the front-fanged *Naja kaouthia* (Family Elapidae; Monocled Cobra) and rear-fanged *Spilotes* (*Pseustes*) *sulphureus* (Family Colubridae; Amazon Puffing Snake) using chromatographic techniques, and toxicity assays were performed to evaluate prey specificity. Despite various 3FTxs being present in abundance within *N. kaouthia* venom, only one 3FTx (alpha-cobratoxin) demonstrated lethal toxicity (<5 μg/g) toward both NSA mice (*Mus musculus*) and House Geckos (*Hemidactylus frenatus*). For *P. sulphureus,* the most abundant 3FTx (sulmotoxin A), a heterodimeric complex, displayed prey-specific toxicity towards House

Geckos, and the second most abundant 3FTx (sulmotoxin B) displayed prey-specific toxicity towards mice. This demonstrates how a relatively simple venom with toxins dominated by one venom protein superfamily (3FTXs) can still allow for the targeting of a diversity of prey.

Venom gland toxin transcriptomes and crude venom transcriptomes were obtained via individual transcripts with 3'RACE (Rapid Amplification of cDNA Ends) and next- generation sequencing to evaluate the abundance, diversity, and molecular evolution of 3FTxs. Venom protein gene expression within rear-fanged snake venom glands revealed trends towards either viper-like expression, dominated by snake venom metalloproteinases, or elapid-like expression, dominated by 3FTxs. For non-conventional 3FTxs transcripts within these glands and within crude venom, approximately 32% of 3FTx amino acid sites were under positive selection, and approximately 20% of sites were functionally critical and conserved.

RNA isolated from crude venom demonstrated to be a successful approach to obtain venom protein transcripts for molecular evolutionary analyses, resulting in a novel approach without the need to sacrifice snakes for tissue. The use of a combined venom gland transcriptome with proteomic approaches aided in characterizing venom composition from previously unstudied rear-fanged snake venoms. This dissertation represents an important step in the incorporation of multiple high-throughput characterization methods and the addition of multiple assays to explore the biological roles of toxins, in particular 3FTxs, within these venoms.

ACKNOWLEDGEMENTS

I would first like to thank my grandparents, especially my grandma, for encouraging and supporting the decisions I have made in my life to achieve a PhD focused on venomous snakes, my childhood dream. Without their financial and emotional support, this achievement would have never been possible. I would next like to thank my parents, they have been understanding about my consistent relocation from state to state and even to the other side of the world so that I could achieve academic and research success. I also appreciate their understanding when at times I was unavailable because of the focus on my degree. I would like to thank all of my family members, my uncles, aunts, cousins, and of course my brother for all their support as well.

I would like to thank my advisor, Steve Mackessy, for guiding me along in this process to obtain my degree. It was invaluable to learn from someone with such an incredible amount of experience in this field and work in a lab that emphasizes venom biochemistry with venomous snake biology. It has always been important for me to work with the animals as well as their venoms.

I would like to thank my biology committee members, Susan Keenan, Mitchell McGlaughlin, and Seth Frietze, for their consistent assistance throughout my PhD, from simple troubleshooting to research execution. They have provided me with the appropriate amount of mental challenges from comprehensive exam

questions to dissertation edits. I would also like to thank other biology faculty at the University of Northern Colorado, especially Judy Leatherman for the guidance she provided during my doctoral supervised teaching. I have learned more about teaching biology from her than any pedagogy class I have taken. I would also like to thank Patrick Burns and Ann Hawkinson for letting me participate in other research projects to broaden my research skillset.

I would like to thank the visiting researchers Ashis Mukherjee and Raquel Sanz-Solar for their assistance and company in the lab. I would also like to thank other research collaborators such as Todd Castoe and Jill Castoe for providing me with resources, training, and science conversations. I would like to thank R.M. Kini for welcoming me into his lab and exposing me to venom research when I was only an undergrad, and for accepting me back as a postdoctoral fellow.

I would like to thank Katelyn Currier, Jessica Rogers, Rebecca Zimmerle, and Julia Torline for all their help in the lab. I can't provide enough thanks to my friends Anthony Saviola, Tony Gandara, Cara Smith, AE Nash, and Brittney Vaughn for being the best lab and office mates ever. Lastly, I would like to thank Tyler Edward Culnan for his support throughout these six years, from moving to Colorado with me to understanding the late nights and weekends I have needed to stay in the lab a little longer.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER I

LITERATURE REVIEW: VENOMS OF COLUBRID SNAKES

Cassandra M. Modahl[a], Anthony J. Saviola[a,b], and Stephen P. Mackessy[a]
[a]School of Biological Sciences, University of Northern Colorado, Greeley, CO, USA
[b]Department of Pharmacology, Weill Medical College of Cornell University, New York, NY, USA

Introduction

Snake venoms represent a critical innovation allowing advanced snakes (Caenophidian) to transition from a mechanical (constriction, as seen in Henophidians) to a chemical (venom) means of subduing prey (Kardong et al. 1997). The complex mixture of proteins and peptides that constitute a snake's venom contribute to multiple biological functions, including immobilizing, dispatching, and digesting prey (Mackessy 2010b). However, rear-fanged venomous snakes are particularly interesting due to the fact that some species of these "colubrid" clades utilize constriction in addition to venom for facilitating prey capture. For example, venom from *Boiga irregularis* (Brown Treesnake) contains a prominent heterodimeric three-finger toxin (3FTx) that is selectively toxic towards lizard and avian prey (Pawlak et al. 2009), and because this 3FTx is nontoxic toward mammalian prey, these snakes will instead constrict mammals (Mackessy et al. 2006). The specific receptor binding exhibited by 3FTxs originates from the accelerated accumulation of nucleotide substitutions within

exons and the resulting changes to protein amino acid sequence and structure (Kini and Doley 2010; Sunagar et al. 2013), and in this case, a taxon-specific toxin has evolved. Venomic techniques combined with transcriptomics, genomics, natural history, behavior and the recognition of venom as a trophic adaptation, offer a powerful approach to unraveling the complex evolutionary history of venoms. This chapter considers the application of combined approaches to study and better understand the venoms of rear-fanged snakes.

Research centered on venom composition and individual protein characterization has provided important insights into the biological roles of venom compounds and the evolutionary relationships of different venomous snakes. This research has also identified compounds which evoke toxic symptoms resulting from snakebite and has directly contributed toward the production of more efficient antivenoms. Because front-fanged venomous snakes belonging to the families Elapidae and Viperidae produce significantly larger venom yields, and are responsible for the vast majority of human envenomations, venom research has primarily focused on species within these two families of snakes (Mackessy 2010b). Rear-fanged venomous snakes, on the other hand, appear to exhibit a less derived venom delivery apparatus, produce significantly lower venom yields, and are generally perceived as non-threatening to humans; as a result, they are generally under-studied relative to the front-fanged snakes (Saviola et al. 2014). The large majority of rear-fanged venomous snakes are unable to deliver lethal quantities of these toxins or even enough toxins to result in systemic envenomations, but at least five species (*Dispholidus typus,*

Thelotornis capensis, Rhabdophis tigrinus, Philodryas olfersii, and *Tachymenis peruviana*) are believed to have caused human fatalities (Kuch and Mebs 2002; Mackessy 2002; Prado-Franceschi and Hyslop 2002). Increasing awareness of severe, at times fatal, envenomations from rear-fanged snakes has led to a slowly growing interest in their venoms. In addition, advances in research techniques have resulted in a modest increase of data on individual toxins and on the composition and complexity of rear-fanged snake venoms.

Even though a single species may produce a venom with more than 100 protein components, snake venom proteins belong to a small number of enzymatic and non-enzymatic superfamilies. Some of these well-recognized venom protein families include phospholipases A_2 (PLA_2s), serine proteinases, snake venom metalloproteinases (SVMPs), three-finger toxins (3FTxs), proteinase inhibitors and lectins (Mackessy 2010b). These major venom protein families are found in almost all snake venoms, including many rear-fanged snake venoms (Mackessy 2002). In general, rear-fanged snake venoms show lower complexity than the venoms of front-fanged snakes, with upwards of 40 expressed proteins visible following 2D SDS-PAGE (two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis), while front-fanged snake venoms show considerably higher complexity. In several cases, rear-fanged snake venoms have also been documented to contain novel protein superfamilies, thus providing a more comprehensive view of venom evolution (Ching et al. 2006; Pawlak et al. 2009; OmPraba et al. 2010). With the introduction of high-throughput proteomic and nucleic acid sequencing methods,

detailed venom descriptive work encompassing the proteome, transcriptome and genome is now possible with the relatively small amount of starting material obtained from rear-fanged snakes. These techniques, combined with protein biochemical characterizations and snake natural histories, will continue to elucidate the evolutionary history and biological roles of rear-fanged snake venom components. The goal of this chapter is to provide a review of previous work, current research and methods, and future applications involving rear-fanged snake venoms.

Previous and Current Research on Rear-Fanged Snake Venoms

Classical Approaches

Since the mid-twentieth century, the study of snake venom toxinology has developed into a formalized scientific discipline. Originally, rear-fanged (opisthoglyphic) snakes were regarded as non-threatening to humans, but the tragic deaths of herpetologists Karl Schmidt (due to envenomation from *Dispholidus typus*) and Robert Mertens (*Thelotornis capensis*) brought attention to the venomous potential of bites from rear-fanged snakes. These events initiated an increase in studies on these venoms (Weinstein et al. 2011), and the discovery that many rear-fanged venom secretions exhibit a complex immuno-identity with numerous medically-important viperid and elapid species (Minton and Weinstein 1987), further stimulated research endeavors into rear-fanged snake venoms. Some lethal venom components, such as 3FTxs, PLA_2 enzymes and SVMPs, that were once thought to be found exclusively in elapid or viperid venoms, now appear to be significantly abundant compounds in numerous

species of rear-fanged snakes (Mackessy 2002; Fry et al. 2003a). However, due to limited accessibility of specimens and low venom yields obtained during extractions, studies involving rear-fanged snake venoms have progressed slowly compared to the extensive work examining front-fanged snake venoms (Mackessy 2002). The now common utilization of anesthetics, such as ketamine hydrochloride, followed by a subcutaneous injection of the parasympathomimetic pilocarpine hydrochloride to stimulate venom secretion has not only improved snake handling and safety for both the animal and handler but has also resulted in greatly increased venom yields (Hill and Mackessy 1997; Mackessy 2002). Now that it is possible to obtain sufficient quantities of venom, coupled with the continuing advancements in biochemical characterization and high-throughput proteomic, transcriptomic and genomic techniques, it is feasible to develop a much broader understanding of the composition and complexity of rear-fanged snake venoms.

Traditional methods, such as the use of one and two-dimensional gel electrophoresis, provide a quick and basic approach for identifying venom compounds present in crude venoms. For 1D SDS-PAGE (one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis), as little as 10-30 µg of crude venom separated in 12% acrylamide precast gels provides a clear molecular fingerprint of potential venom compounds, allowing for inter- and intraspecific comparisons of venom variation (Mackessy et al. 2006; Peichoto et al. 2012). Rear-fanged venoms that have been studied typically demonstrate a greater complexity in the higher molecular mass regions following SDS-PAGE

(Peichoto et al. 2012). Two-dimensional gel electrophoresis provided additional venom compositional information in an analysis of South and North American opisthoglyphous snake species, allowing for the detection of multiple protein isoforms that shared similar molecular masses but differed in isoelectric points. Acidic proteins in the 30-40 kDa range, which would have been difficult to distinguish using only one-dimensional gel electrophoresis, showed differential expression in *Philodryas* sp. and for *Trimorphodon biscutatus lambda* (Peichoto et al. 2012). The greater resolution provided by 2D gel electrophoresis is also an ideal fractionation method for in-gel trypsin digestion and mass spectrometry (MS) analysis. This technique allows for the identification of multiple protein isoforms that can exist in a venom protein superfamily, such as the multiple matrix metalloproteinase isoforms found in the rear-fanged snake *Thamnodynastes strigatus* (Ching et al. 2012).

Whole organism toxicity, another traditional venom characterization method, allows for identification of lethal doses (LD_{50}); this dosage reflects the amount of a substance required to kill half of the injected organisms within a 24-hour period. Low LD_{50} values indicate the presence of potent (often neurotoxic) venom components, but it can be difficult to obtain enough material, especially purified individual venom components, from some rear-fanged snakes (Mackessy 2002). LD_{50} values have classically been determined using a mouse model because mice are easy to maintain, can be obtained in an array of essentially "reagent grade" strains, and are included in the diet of many venomous snakes (da Silva and Aird 2001). Because the diets of rear-fanged venomous snakes

often encompass a broader range of prey taxa, species that regularly feed on non-mammalian prey may produce venom toxicity and other physiological data in mice that are not biologically relevant (Pawlak et al. 2006; Pawlak et al. 2009). For example, NSA mice showed no adverse effects from the 3FTx irditoxin (from *B. irregularis* venom) at doses of at least 25 µg/g, whereas House Geckos (*Hemidactylus frenatus*) and domestic chickens (*Gallus domesticus*) exhibited rapid flaccid paralysis, dyspnea and increased respiratory rates at all doses tested, with an LD_{50} <0.55 µg/g (Pawlak et al. 2009). This correlates closely with the diet of *B. irregularis*, which frequently feeds on birds and lizards, and demonstrates the importance of venom as a tropic adaptation and the need to acknowledge snake natural history when elucidating venom protein biological roles. LD_{50} values from mice alone (crude venom - 18-31 µg/g) (Mackessy et al. 2006), would not have revealed the complexities of *B. irregularis* venom and would not have detected the presence of a prey-specific toxin.

Analyses of rear-fanged venoms by HPLC (high performance liquid chromatography) size exclusion chromatography has revealed the presence of larger mass proteins, with acetylcholinesterase and metalloproteinase activities limited to the first peaks, cysteine-rich secretory proteins (CRiSPs) found in the second peaks, followed by PLA_2s, and then 3FTxs when present (Peichoto et al. 2012). Ion exchange chromatography, especially cation exchange, has been shown to be an effective purification first step for 3FTxs present in rear-fanged snake venoms and was used successfully for the isolation of the 3FTxs fulgimotoxin and denmotoxin (Pawlak et al. 2006; Heyborne and Mackessy

2013). Anion exchange columns have been successful for purifying SVMPs from rear-fanged snake venoms (Weldon and Mackessy 2012). Reversed-phase (RP) HPLC is a common final polishing step for the removal of salts from size exclusion or ion exchange chromatography, but this method can result in denaturation of some venom proteins, in particular metalloproteinases and other enzymes.

RP-HPLC has also been utilized as a first step for descriptive venomics because this technique provides a clear image of crude venom complexity by separating protein isoforms and exhibiting relative abundance of venom protein superfamilies when combined with SDS-PAGE or MS (Fry et al. 2003c; Pawlak et al. 2006; Calvete et al. 2009). A combination of liquid chromatography and soft ionization mass spectrometry (LC/MS) has been used to analyze crude rear-fanged snake venoms, including species from Colubrinae, Homalopsinae, Natricinae, Psammophiinae, Pseudoxyrhophiinae, and Xenodontinae (Fry et al. 2003c). An advantage to this technique is that it can be performed with limited amounts of material. MS molecular masses and LC retention information can also provide an idea of represented venom protein superfamilies in a crude venom (Fry et al. 2003c). However, ion suppression with co-eluting proteins is a problem with electrospray mass spectrometry (ESI-MS), and proteins of lower abundance can be overlooked, resulting in missing venom peptide information and incomplete proteomic characterization.

For determining the amino acid sequences of purified venom proteins, N-terminal sequencing (Edman degradation) has been broadly used. N-terminal

sequencing and tandem MS for *de novo* sequencing can provide reliable amino acid sequences, and automated *de novo* sequencing tools are increasingly becoming more sensitive and widely used for determining the amino acid sequences of complex mixtures of proteins. However, identification of proteins from rear-fanged snake venoms can still be problematic given the limited amount of database information available for rear-fanged venom protein sequences. A post-synaptic neurotoxin was isolated in the rear-fanged *Rhamphiophis oxyrhynchus* (Rufous Beaked Snake), but lacked sequence homology to any previously identified snake venom toxin in any database, making it difficult to determine what venom protein family this neurotoxin represented (Lumsden et al. 2007).

Venoms are composed of both enzymatic and non-enzymatic proteins, as well as small peptides and other organic molecules (Mackessy 2010b), and numerous enzyme assays have been developed for the detection of the major snake venom enzyme superfamilies. These assays include substrates that can indicate the presence of proteases (SVMPs and serine proteinases), acetylcholinesterases, PLA_{2s}, L-amino acid oxidases, hyaluronidases, and phosphodiesterases in rear-fanged snake venoms (Mackessy 2002). Proteolytic activity has been assayed for using several substrates, including casein yellow, azocasein, collagen, and fibrinogen (Sanchez et al. 2014). Zymogram gels, which are copolymerized with gelatin, have also been used to characterize rear-fanged snake venom proteins with proteolytic activity (general endoproteinase activity) (Hill and Mackessy 2000; Weldon and Mackessy 2010).

Using azocasein substrate, SVMP activity has been identified in many rear-fanged snake venoms, including the venoms of *Dispholidus typus*, *Philodryas sp., Hydrodynastes gigas, Hypsiglena torquata* and *Alsophis portoricensis* (Mackessy 2002; Peichoto et al. 2007; Weldon and Mackessy 2012). This list of species includes both New World and Old World rear-fanged snakes and is suggestive of potential local tissue damage and hemorrhage if bitten by these species (Peichoto et al. 2012; Sanchez et al. 2014). Currently there are several rear-fanged SVMPs that have been further characterized, such as patagonfibrase from *Philodryas patagoniensis* and alsophinase from *Alsophis portoricensis*, both of which demonstrate alpha-fibrinogenolytic and hemorrhagic activities (Peichoto et al. 2007; Weldon and Mackessy 2012). *Philodryas patagoniensis* venom has been reported to contain proteolytic activity greater than the venom of *Bothrops alternatus*, and the venom of *P. baroni* was reported to exhibit proteolytic activity 25 times greater than the activity reported for *B. jararaca* (Sanchez et al. 2014). Hemorrhagic SVMPs and serine proteinases are responsible for severe local inflammation and tissue necrosis in human envenomations, and significant bleeding has been reported from rear-fanged snake envenomations, likely due to the presence of these toxins (Weinstein et al. 2011). Assaying rear-fanged snake venoms for proteolytic activity, particularly SVMP activity, can be useful to predict the potential envenomation hazard these snakes could pose to humans.

Snake venom metalloproteinase classes differ in structure with regard to domain composition; P-Is have only the metalloproteinase domain, P-IIs have an

additional disintegrin domain, and class P-IIIa-c have a metalloproteinase, disintegrin, and cysteine-rich domain, with P-IIId having an additional lectin domain (Mackessy 2010b). The only SVMPs to date that have been discovered in rear-fanged snake venoms have been of the P-III class, which have been characterized in several venoms, including *Dispholidus typus* and *Alsophis portoricensis* venoms, among others. Although full venomic analyses (protein digestion, followed by peptide mass fingerprinting) was not utilized to identify protein families in several of these studies, SVMP activity was detected using an azocasein substrate confirming the presence of SVMPs in these venoms. A combined proteomic and transcriptomic analyses of the venom of *Philodryas olfersii,* a rear-fanged venomous snake of South America with growing medical significance, revealed toxin similarities to those of snakes belonging to the family Viperidae, with the P-III class of SVMPs being the most abundant protein in the venom (Ching et al. 2006). P-III SVMPs are also the most abundant compounds in the venoms of *Thamnodynastes strigatus* (Ching et al. 2012) and of *Hypsiglena* sp. (McGivern et al. 2014). Both one and two dimensional gel electrophoresis further confirmed the presence of P-III SVMPs not only in *P. olfersii,* but also in *P. patagoniensis, P. baroni,* and *Hypsiglena torquata texana* venoms (Peichoto et al. 2012). It is thought that during the evolution of a front-fanged venom system, the various domains observed in P-III SVMP were gradually lost. The P-I and P-II classes of SVMPs are currently only found in Elapidae and Viperidae venoms (Mackessy 2010b).

Acetylcholinesterase activity has been reported in several rear-fanged snake venoms, with this activity being most prominent in venoms of *Boiga* species such as *B. irregularis* (Mackessy 2002). This acetylcholinesterase activity appears to be substrate-specific as it lacks activity toward a butyrylcholine substrate (Mackessy 2002). Acetylcholinesterase activity is commonly detected in venoms with the use of the substrate acetylthiocholine that reacts with dithiobisnitrobenzoate to produce a colorimetric determination of activity. For the detection of PLA_2 activity, 4-nitro-3-(octanoyloxy) benzoic acid and egg yolk phosphatidycholine Type IV substrates have been used to assay activity in venoms of *Boiga dendrophila, Diadophis punctatus regalis, Dispholidus typus, Leptodeira annulata, Malpolon monspessulanus, Rhabdophis subminiata, Thelotornis capensis, Rhamphiophis oxyrhynchus,* and *Trimophodon biscutatus lambda* (Hill and Mackessy 2000; Huang and Mackessy 2004). The PLA_2 trimorphin was been purified and characterized from *Trimorphodon biscutatus lambda* venom.

CRiSPs are also widespread in reptile venoms and exhibit a remarkable degree of sequence conservation (Peichoto et al. 2009; Mackessy and Heyborne 2010), and several members of this superfamily have been found to interact with different target proteins, such as cyclic nucleotide-gated ion channels as well as L-type Ca^{2+} and K^+ channels (Yamazaki and Morita 2004). The biological functions of many CRiSPs remain relatively unknown. A CRiSP isolated from the rear-fanged snake *Helicops angulatus* has been shown to exhibit robust neurotoxic activity that results in immediate respiratory paralysis in mice (Estrella

et al. 2011), while patagonin, a CRiSP characterized from the venom of *P. patagoniensis,* was found to cause muscular damage (Peichoto et al. 2009). Rear-fanged venom CRiSPs appear to show the same conservation of structure and diversification of function as seen for 3FTxs.

In the absence of biochemical and biological assays, it can be difficult to predict the activity of venom proteins. Large venom protein superfamilies such as 3FTxs and PLA_2s can exhibit a diversity of activities, ranging from neurotoxicity as a result of specific receptor binding to general cytotoxicity resulting in tissue necrosis. Researchers should be careful assigning protein activity based solely upon sequence similarity to other proteins or if only limited biochemical assays have been conducted. In the case of PLA_2s, pharmacological effects may be dependent or independent of enzymatic activity; therefore, a biochemical assay focused only on PLA_2 enzymatic activity could overlook other pharmacological activities, such as neurotoxicity (Mackessy 2010b). Many high-throughput venom descriptive techniques, such as those based on MS/MS (tandem mass spectrometry) data, are limited when it comes to evaluating of structure-function variation within venom protein families. Venom proteins can share similar amino acid sequences and have different structural arrangements as a result of posttranslational modifications or interactions between other venom proteins or substrates. It is also possible for toxins to have similar structural appearances but exhibit vastly different receptor targets or activities. An example of this is the diversity of biological activities exhibited by venom 3FTxs, which include

neurotoxicity, enzyme inhibition, cardiotoxicity, cytotoxicity, ion channel blockage and anticoagulation effects (Kini and Doley 2010).

Venoms of several rear-fanged snakes in the family Colubridae (*sensu stricto*) contain 3FTxs that maintain the same conserved three β-sheet stabilized loops (from which the name "three-finger toxin" originated) commonly seen in 3FTxs from elapid venoms. Several 3FTxs from rear-fanged snakes have taxon-specific receptor binding affinities which has not been observed for elapid 3FTxs. Denmotoxin, from venom of the rear-fanged *Boiga dendrophila* (Mangrove Catsnake), was the first prey-specific 3FTx identified and displayed potent postsynaptic neuromuscular activity by irreversibly inhibiting chick biventer cervicis nerve-muscle preparation twitches, but it induced much smaller and reversible inhibition of twitches in mouse hemidiaphragm nerve-muscle preparations, suggestive of a bird-specific postsynaptic affinity (Pawlak et al. 2006). Irditoxin, a lizard and avian specific 3FTx from *B. irregularis*, was identified shortly after denmotoxin (Pawlak et al. 2009), and recently another prey-specific 3FTx, fulgimotoxin, was discovered in a New World rear-fanged snake, *Oxybelis fulgidus*, indicating that this phenomenon is not limited to Old World species and is likely more common in rear-fanged snake venoms (Heyborne and Mackessy 2013). Based on 1D SDS-PAGE and other data, 3FTxs are present in numerous venoms from rear-fanged snakes (Saviola et al. 2014).

More Recent Approaches

"First generation" venomics has been an exceptionally successful means to generate near-complete catalogs of venom proteins, and this approach has

also been applied to venoms of rear-fanged snakes (Calvete et al. 2009). In recent years, the emergence of “omic” technologies has revolutionized venom research by integrating detailed high-throughput approaches to generate systematic venom studies involving whole genomes, transcriptomes and proteomes (Calvete 2013). To date, a comprehensive approach, with (proteomics) MS/MS peptide sequencing of separated venom components (usually by RP-HPLC or 2D gel electrophoresis) combined with a species-specific venom gland transcriptome, has provided the most complete venom compositional coverage (Sunagar et al. 2014). MS/MS identification of peptide sequences relying on available protein sequence databases, such as the Mascot online server, can overlook unique isoform variations and present a limited capacity to identify novel venom protein if only small peptide fragments are used for protein identification. By generating a complementary transcriptome, MS/MS peptide sequences can be more precisely identified to the corresponding transcript, and translated transcripts will provide full protein sequences. Obtaining full sequences using only proteomic methodologies (such as N-terminal sequencing and MS/MS *de novo* sequence determinations from many peptide fragments) would otherwise be much more labor-intensive and expensive.

Approaches to venom characterization have largely focused on mass spectrometry to generate complete venom profiles. The two primary MS methods for whole proteins include matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) MS and electrospray ionization MS (Kukhtina et al. 2000). These methods are frequently used to provide more accurate molecular masses

for individual venom components and peptide fragments, and both allow for high-throughput analysis of complex samples. Mass spectrometric *de novo* sequence determination is especially of interest for protein sequences that have a blocked N-terminus, making it more difficult to determine the amino acid sequence from Edman degradation. Rear-fanged snake venom 3FTxs commonly have an N-terminal pyroglutamate which must be removed prior to Edman sequencing (Pawlak et al. 2009; Heyborne and Mackessy 2013).

Top-down and bottom-up approaches are seen in proteomic literature regarding rear-fanged venomous snakes, where a top-down approach is done with intact venom proteins and a bottom-up approach is accomplished using proteolytic peptide mixtures. A top-down MALDI-TOF-MS method using rear-fanged snake venom has revealed as many as forty-nine distinct protein masses (Peichoto et al. 2012). Top-down strategies allow for more complete characterization of protein isoforms and post-translational modifications (Han et al. 2008). Post-translational modifications found in rear-fanged snake venom proteins have yet to be studied in detail, and many opportunities exist for continued work using top-down MS methods.

A bottom-up approach, such as tandem MS performed on proteins digested with proteases such as trypsin (most commonly used), chymotrypsin or Glu-C, generates a spectrum of fragmented singly charged peptide ions that can be matched to databases for protein identification (peptide mass fingerprinting) or can be used for *de novo* sequence determination (Chapeaurouge et al. 2015). Collision-induced dissociation (CID) is the most widely used MS/MS technique

for this type of venom analysis. This technique creates a series of backbone fragmentations at the peptide bond, resulting in b- and y-fragment ions. MASCOT, SEQUEST, or other databases are searched using algorithmic comparisons of proteins derived from genomic sequencing or known protein amino acid sequences to identify unknown proteins based on their peptide fragment spectra.

However, post-translational modifications of venom proteins are not detectable when examining the venom genome or transcriptome, and discrepancies between the proteome and the transcriptome of a single species have been noted (Pahari et al. 2007; Sunagar et al. 2014). Translation blockages are also not detectable based on transcriptome data alone (Wang et al. 2010), and therefore genomic and transcriptomic data do not fully represent the compounds that may constitute a species' venom. In addition, the increased sensitivity of transcriptomics results in all venom gland mRNA (messenger ribonucleic acid) being sequenced, and therefore it can be difficult to discern which transcript sequences are translated and secreted as venom proteins and which are simply endogenous cellular proteins. Adopting a combination of 'omic' approaches allows for a species-specific transcriptome database of all potential venom protein components in a venom to be matched with MS/MS-generated peptide fragments from the crude venom proteome. Proteomics can also be used to check the accuracy of transcriptome assembly and translation. If peptide MS/MS sequences do not match a single species-specific transcriptomes, it could be suggestive of erroneous contig (contiguous sequence from overlapping

DNA reads representing a transcript) assembly, sequencing errors that resulted in a reading frame shift, or incorrect reading frame selection (Calvete 2014). Therefore, a combination of genomic, transcriptomic and proteomic data is necessary to fully understand venom composition and evolution.

Currently, several rear-fanged genera, including *Boiga*, *Hypsiglena* and *Philodryas*, are the best characterized venoms using a combination of proteomic (general venom description and biochemical activity) and transcriptomic data. For *P. baroni, P. olfersii,* and *P. patagoniensis,* it has been shown that the majority of venom proteolytic activity is from SVMPs, with low levels of activity towards substrates for serine proteinases (Sanchez et al. 2014). The venom of *Boiga irregularis* was found to be dominated by 3FTxs, while that of *Hypsiglena* was dominated by SVMPs; the venom of *Boiga* appeared more “elapid-like”, while the venom of *Hypsiglena* was more “viper-like” (McGivern et al. 2014). A combination of proteomics and transcriptomics is ideal for rear-fanged venomous snakes, particularly considering that these venoms are not well characterized and have been observed to have novel protein superfamilies that would be missed using MS/MS peptide matching techniques alone (due to the lack of rear-fanged snake venom protein sequences in current databases). Genomic and transcriptomic data is becoming more readily available for venomous snakes (Rokyta et al. 2012b; Rokyta et al. 2013; Vonk et al. 2013) and is enhancing the understanding of evolutionary relationships between venom compounds, and the snakes which produce them. As these databases grow, they will allow for the investigation of the multiple levels of transcriptional and translational regulation of

venom proteins that give rise to the variation that is seen in venom composition between species and even within individuals.

Venom gland transcriptome analyses are powerful for determining venom transcript expression, but genomic sequences can provide insight into venom gene transcriptional regulation (i.e. promoter sequences) and mechanisms resulting in venom protein diversity (i.e. alternative splicing events and/or gene dosage effects). A combined genomic and transcriptomic approach will also allow for splicing variations to be identified. Splicing variations allow for different functional proteins to be transcribed using the same exons, and this may help lead to binding to different receptors. New exons may also be inserted into the gene. Three-finger toxins in rear-fanged snakes have been found to have extended N-terminal segments compared to elapid and viperid 3FTxs, and for denmotoxin, this is the result of a newly inserted exon two. Currently, the function of this longer N-terminal region is unclear, but this is currently the only full 3FTx gene sequence for a rear-fanged snake and it provides insight into additional mechanisms of evolution of these toxins (Pawlak and Kini 2008).

High-throughput Proteomic Approaches to the Study of Snake Venoms

Protein chemistry methodologies have been utilized to examine snake venoms very early in the history of modern venom biochemical research. However, recent advances in proteomic techniques and the utilization of mass spectrometry have greatly expanded the understanding of venom composition, allowing for the field of venom proteomics (venomics) to flourish. The term ‘snake

venomics' was developed by Calvete and coworkers and has been important as a standardized venom characterization protocol and a semi-quantitative estimation of venom protein relative abundances (Calvete 2013). By measuring the absorbance at 215 nm during the primary RP-HPLC separation (which roughly correlates with the abundance of peptide bonds), a percentage value can be assigned to each chromatographic peak, and in combination with a densitometric lane scan of SDS-PAGE run using individual chromatographic peaks, the relative percentage of different venom components that make up the overall crude venom can be determined (Calvete et al. 2009). Standard venomic protocols involve chromatographic (usually RP-HPLC) and electrophoretic techniques to separate crude venom proteins which are then digested into peptides with proteolytic enzymes (most commonly trypsin). Individual peptide ions can also be fragmented by collision-induced dissociation, with the resulting daughter fragment ions identified by manual inspection. This approach identifies venom proteins based upon multiple lines of evidence, including molecular mass, peptide mass matching, several peptide sequence identifications, and determination of the number of cysteine residues present. These data typically allow for the identification of most toxin classes found in snake venoms (Calvete 2014). These methodologies, incorporated with biochemical and toxicological data, have allowed for a detailed examination of intraspecific, geographic, and ontogenetic venom variability primarily aimed at addressing the venom composition of dangerously toxic snakes of the families Elapidae and Viperidae (Calvete et al. 2009; Calvete 2014). Venomics also allows for identification of

venom compounds that may be further examined for potential therapeutic value. Although the vast majority of venomic studies have included species that are of great medical significance, emerging studies on the venomics of rear-fanged snakes are elucidating the biological complexities of this poorly studied class of snakes.

Currently, only a few complete rear-fanged venomous snake proteomes are available, and where most commonly a bottom-up strategy has been applied. There are primarily two bottom-up proteomic workflows. There is a “sort-then-break” approach, which includes performing protein fractionation and separation prior to protein digestion, followed by peptide analysis by peptide mass fingerprinting or *de novo* peptide sequence determination (Han et al. 2008). This workflow is seen in the venomic approach to venom profiling as mentioned above and was utilized with the venom of the rear-fanged snake *Thamnodynastes strigatus*. 2D gel electrophoresis was the method of separation before in-gel trypsin digestions and identification of individual protein spots using a MALDI Q-TOF (matrix-assisted laser desorption ionization quadrupole-time-of-flight) Premier mass spectrometer (Ching et al. 2012). Also, several protein SDS-PAGE bands from the venoms of *Trimorphodon biscutatus lambda, Philodryas olfersii, Philodryas patagoniensis, Philodryas baroni,* and *Hypsiglena torquata texana* were digested with trypsin and analyzed with MALDI-TOF/TOF (tandem matrix-assisted laser desorption ionization time-of-flight) to confirm the presence of PLA_2s, CRiSPs and 3FTxs within some of these venoms (Peichoto et al. 2012).

An alternative is the “break-then-sort” approach, where protein digestion is performed without any pre-fractionation/separation and peptides are separated by multi-dimensional chromatography followed by tandem MS analysis (Han et al. 2008). This technique is referred to as “shotgun proteomics”. Both methods are heavily reliant on the high-throughput advances in mass spectrometry, allowing for the identification of multiple peptide fragments to assemble an overall complete venom profile.

“Shotgun” methods involve the production of small sequence fragments of a greater whole that are identified and then assembled into a larger picture. In the case of shotgun venom proteomics, overall venom composition is determined from the identity of the fragmented peptide ions after a whole venom protein digestion. This approach was used to resolve the proteome of the rear-fanged *Cerberus rynchops* (Dog-faced Water Snake) and resulted in the identification of a novel snake venom protein family (veficolins) that is suggested to induce platelet aggregation and/or initiate complement activation (OmPraba et al. 2010). Shotgun proteomics offers an alternative to venomics and can be particularly useful for rear-fanged snake venoms because their venoms are typically much less complex than Elapidae or Viperidae venoms. A shotgun technique can also be used on individual venom proteins; once a venom protein is purified, it can be digested with several different proteases and the resulting peptide fragments from different digestion libraries can be assembled to resolve the complete amino acid sequence (Bandeira et al. 2007).

Shotgun proteomics and venomics both offer insight into the complete composition of rear-fanged snake venoms and are high-throughput and sensitive techniques that can be done using relatively little starting material. These methods can also allow for rapid *de novo* elucidation of primary structure (amino acid sequence and post-translational modifications) of single peptides in a complex mixture or peptides derived by in-solution or in-gel proteolysis of larger proteins. Relative abundances of venom protein families can be estimated using these techniques; however, it is more difficult to determine relative abundances with shotgun venom proteomics. Shotgun proteomics results can be strongly biased, with portions of abundant proteins being overrepresented in many spectra and low abundance protein spectra not being seen at all (Bandeira et al. 2007; Calvete 2014). The increased complexity of the generated peptide mixture requires highly sensitive and efficient separation. On the other hand, with venomics, it is also possible that during RP-HPLC separation, before electrophoresis and MS/MS sequencing, highly hydrophobic and/or large proteins may elute poorly and be absent or underrepresented in abundance (based upon chromatographic peaks and electrophoretic results).

Both of the above methods rely on the identification of peptide masses/ionization patterns. Peptide mass fingerprinting and *de novo* MS/MS sequence determination methods are high-throughput, less labor intensive, and more-cost effective than N-terminal sequencing, but a limitation to the reliance on peptide masses is that certain combinations of amino acids can have indistinguishable masses, therefore creating ambiguity. An example of this is the

assignment of isobaric (Ile/Leu) or quasi-isobaric residues (Lys/Gln or Phe/Met-ox), although methods such as high-resolution Fourier-transform ion cyclotron resonance or Orbitrap mass analyzers can be used to discriminate between quasi-isobaric residues (Calvete 2013). It can also be difficult to identify correctly all peptide sequences with peptide mass fingerprinting, especially those with unexpected modifications or from proteins that are absent from databases. Identification based on shared peptide sequences in databases often does not allow differentiation between isoforms, and snake venoms can have multiple different isoforms present, each with potentially different pharmacological activities (Kini and Doley 2010; Mackessy 2010b; Calvete 2014). Complete amino acid sequences for large, unknown proteins from bottom-up methods is also not possible due to the incomplete recovery of a full tryptic peptide set. Although tryptic digestion followed by LC-CID-MS/MS (liquid chromatography collision-induced dissociation tandem mass spectrometry) is ideal for the identification of toxin classes, it does not provide information about the quaternary structure of individual toxins or toxin activities. After individual venom proteins are cleaved into peptide fragments, they can no longer be used for follow-up biochemical or pharmacological assays. In recent years, venomics and shotgun venom proteomics have become highly sensitive techniques to provide information obtaining to overall venom composition for a snake species, but biochemical and pharmacological assays are needed for complete venom protein characterization.

High-throughput Approaches to the Study of Snake Venoms

The majority of colubrid venom studies have focused on the protein composition and enzymatic properties of these venoms, with relatively few published venom gland transcriptomes or venom protein transcripts. Although venom compositional and biochemical studies can help to infer clinical symptoms of envenomation and the biological roles of these proteins, venom protein transcripts can also be used to assess venom composition and to predict protein activity. Transcriptomic studies can provide a starting point for proteomic methods when crude venom material is lacking or of low venom yield, which is commonly an issue when working with rear-fanged venomous snakes. A venom protein-coding transcript can be translated to acquire an entire protein amino acid sequence and this allows for identification of protein superfamilies and functional protein domains. Obtaining the sequences of toxin mRNAs can also assist in the assembly and completion of protein sequences where trypsin digests or N-terminal sequencing provides only partial sequence. Transcripts also provide information about the evolutionary history of venom protein superfamilies and can be used for the reconstruction of ancestral sequences. One can then explore questions such as the origin of venom and the mechanisms responsible for venom evolution and adaptation (Casewell et al. 2012; Casewell et al. 2013). Transcript sequences are needed for positive selection analysis within protein superfamilies, because to establish protein amino acid sites under positive selection, protein-coding mRNA transcripts must be used to determine where

single-nucleotide polymorphisms occur and whether there are nonsynonymous mutations (Sunagar et al. 2013).

Transcriptomic and genomic methods offer many exciting opportunities for future studies, and the cost of next-generation DNA sequencing is becoming more affordable. With rear-fanged snakes comprising several Colubroidea families and subfamilies, many novel venom transcripts likely exist and remain unexplored. Colubrid transcriptomes and genomes offer the opportunity to identify novel venom protein families, scaffolds, and provide insight into the evolutionary histories of ubiquitous venom protein families (Ching et al. 2006; OmPraba et al. 2010; Fry et al. 2012).

Currently, complete venom gland transcriptomes have only been published for *Philodryas olfersii* (Ching et al. 2006)*, Cerberus rynchops* (OmPraba et al. 2010)*, Thamnodynastes strigatus* (Ching et al. 2012)*,* and *Boiga irregularis* and *Hypsiglena* sp. (McGivern et al. 2014), and some venom transcript sequences from *Dispholidus typus, Telescopus dhara, Trimorphodon biscutatus, Liophis miliaris, Liophis poecilogyrus, Leioheterodon madagascarensis, Psammophis mossambicus,* and *Rhabdophis tigrinus* are also available (Fry et al. 2012). The venom gland transcriptome from *Cerberus rynchops* revealed a novel venom protein family, ryncolin, that was the first discovered venom protein to exhibit sequence similarity to ficolin (a mammalian protein with collagen-like and fibrinogen-like domains) (OmPraba et al. 2010). The venom gland transcriptome from *Thamnodynastes strigatus* was found to be largely composed of matrix metalloproteinases, unrelated to the metalloproteinases found in other

Colubroidea snake families (Ching et al. 2012). This was the first time that matrix metalloproteinases were identified as a prominent venom component and were discovered to make up the majority of *T. strigatus* venom (both in abundance of transcripts and proteins) (Ching et al. 2012). A combined RNA-seq (ribonucleic acid sequencing, completed with the generation of complementary deoxyribonucleic acid libraries [cDNA]) and mass spectrometry analysis of venom glands and venoms from two species indicated that there are very different venom compositional "strategies" present among rear-fanged snakes, reminiscent of differences seen between elapid and viperid species (McGivern et al. 2014). Other protein families identified in an analysis of several rear-fanged snake venom glands include lipocalin, phospholipase A_2 (type IIE), vitelline membrane outer layer protein and ribonucleases (Fry et al. 2012). Besides identifying novel venom protein families, rear-fanged snake venom gland transcriptomes have provided sequence resources that have deconstructed venom protein evolutionary histories, such as the evolution of C-type natriuretic peptides throughout Colubroidae (Ching et al. 2006).

The number of rear-fanged snake venom transcripts and genes will certainly increase as sequencing technologies have become more available and affordable. Next-generation sequencing (NGS) allows for multiple venom gland transcriptomes to be sequenced in parallel and removes the need for tedious *Escherichia coli* cloning procedures. The majority of current rear-fanged snake venom gland transcriptomes have been constructed by first selecting mRNA from gland tissue (usually from a gland removed three to four days after venom

extraction), generating cDNA libraries by reverse transcription, cloning these sequences with the use of plasmid vectors and transformed *Escherichia coli*, and then randomly picking clones to be sequenced with chain-terminating Sanger sequencing technology. This methodology can introduce bias into a study since smaller cDNA fragments have a higher transformation efficiency, or transcripts could be partially expressed in *E. coli* with lethal effects (Durban et al. 2011). There is also a high likelihood of missing many mRNAs that occur in low abundance if not enough colonies are selected. NGS techniques remove these biases and create larger sequence assemblies.

Given that there is extensive proteomic work remaining to determine the protein families that compose rear-fanged snake venoms, care must be taken when identifying “venom protein” transcripts from a rear-fanged snake venom gland transcriptome without proteomic evidence. It is likely that transcripts that encode endogenous cellular proteins that are not found in the venom will be sequenced through a transcriptomic approach. Several published rear-fanged snake transcriptomes have provided proteomic evidence to support the translation and secretion of identified venom protein transcripts with the use of two-dimensional gel electrophoresis and/or HPLC separation and mass spectrometric analysis (Ching et al. 2006; OmPraba et al. 2010; Ching et al. 2012), and until there is a better understanding of the venom protein families occurring in rear-fanged snake venoms, applying a combined transcriptomic and proteomic approach would be the best way to decipher venom protein composition in this understudied clade of snakes.

Ancestral venom proteins have diverse activities and perform a variety of physiological roles for the snake in a diverse set of its tissues (Fry 2005). Recent analyses have determined that toxin homologs are expressed in other tissues, suggestive that these “toxins” are either co-expressed in many tissues or are “reverse recruited” from the venom gland for other physiological roles in other tissues (Casewell et al. 2012; Hargreaves et al. 2014). It has been suggested that these genes are co-expressed in a variety of tissues and then following gene duplications, are restricted to expression in the venom gland after transcriptional regulation changes within other tissues. Therefore, having multiple transcriptomes from different tissues is needed to compare the expression patterns of related genes and to better understand the evolutionary history of venom gene superfamilies and the events following venom gene duplication, subfunctionalization, and neofunctionalization within these gene families, as well as understanding how these processes influences venom protein adaptability. Future tissue transcriptomes will also help to provide insight into the origin and evolution of venom proteins, because non-toxin homologues are also needed to construct venom gene trees (Casewell et al. 2012). Furthermore, there is a need to compare different tissues and to determine differential expression of venom gene homologs in other tissues and the possibility of “reverse recruitment.” Current studies are limited in that only a handful of venom gene families, tissues and snake species have been analyzed (Casewell et al. 2012; Hargreaves et al. 2014). Venom evolution is a very complex and dynamic system, and research in

this area has applications for studies involving other proteins that experience accelerated evolution and novel functionality gain.

The mechanisms underlying the regulation of venom gene expression remain largely unstudied. An understanding of the gene regulatory elements such as promoters and enhancers as well as the transcription factors that bind to these regions, including related epigenetic mechanisms (DNA methylation, etc.), remain largely unexplored in venomous snakes. Techniques such as ChIP-Seq (Chromatin Immuno-Precipitation-sequencing) and RIP-seq (RNA Immuno-Precipitation-sequencing) are possible future approaches to determine the regulatory proteins binding to DNA and RNA involved in the transcription and translation of venom genes and transcripts.

Changes to DNA sequence can directly affect gene products and influence the evolution of a protein. Although protein sequences can provide information regarding potential structure and function, the coding gene sequence (CDS) can reveal hidden single nucleotide polymorphisms (SNPs) or elucidate molecular evolutionary patterns. Venom gene nucleotide polymorphisms can have significant impacts on venom proteins; a dinucleotide deletion in a 3FTx gene resulted in the loss of neurotoxic activity in *Aipysurus eydouxii* (Marbled Sea Snake) venom (Li et al. 2005b, a).

Venom genes experience increased nucleotide substitution rates, especially within exon regions, as compared to other protein-coding genes (Doley et al. 2009). Venom protein genes have the flexibility to accumulate mutations at an increased rate due to the presence of multiple gene copies

resulting from gene duplications and subfunctionalization. If one sequence develops a detrimental mutation, other copies of the gene remain present and functional. This mechanism also allows for venom gene neofunctionalization. Venom multigene families have been identified as evolving by this 'birth-and-death' gene model (Fry et al. 2003b; Vonk et al. 2013). Venom proteins typically possess a stable structural core maintained by multiple disulfide bonds, and non-synonymous nucleotide substitutions that alter non-structural, surface-exposed residues can change protein-targeting interactions, such as targeting new receptor types (Doley et al. 2009). This allows for a venom protein family to develop multiple activities, and having multiple gene products provides a selective advantage over the optimization of a single gene product by allowing for evolutionary "experimentation".

There have been several mechanisms proposed to explain accelerated gene neofunctionalization rates within venom gene superfamilies, including accelerated segment switching in exons to alter targeting (Doley et al. 2009), and rapid accumulations of point mutations in exposed residues (Doley et al. 2009; Sunagar et al. 2013). Transposable elements have also been shown to produce protein diversity and may play an important role in the evolution of species-specific snake venom genomes and venom compositions. The python genome was found to have a large abundance of retroelements in comparison to its size (Castoe et al. 2013), especially LINEs (long interspersed elements). LINEs have been associated with creating protein diversity by carrying along genetic material from transposition events and therefore resulting in additional exon segments.

Transposable elements can contribute to non-homologous recombination, one of the mechanisms responsible for generating gene duplications (Zhang 2003; Hargreaves et al. 2014).

Recently, the first snake genomes, *Python molurus bivittatus* (Burmese Python) and *Ophiophagus hannah* (King Cobra) were completed (Castoe et al. 2013; Vonk et al. 2013). These snake genomes provide genomic scaffolds that will facilitate assemblies and annotation for other snake genomes. Genomic sequencing also provides potential full venom protein sequences. Databases with genomic sequences can be translated into all possible reading frames and matched to resulting peptide fragments obtained by tandem mass spectrometry. Many proteomic techniques will therefore be aided by the addition of complete snake genomes and in combination with venom gland transcriptomes, these genomes will provide overall insight into venom protein expression and evolution. A complete genome of a rear-fanged venomous snake is currently unavailable, but once finished, this genomic information will be accessible to compare venom multigene families within different snake families. For example, in the case of the rear-fanged snake *Boiga dendrophila*, the full gene sequence of the venom 3FTx denmotoxin was found to exhibit unique gene organization compared to 3FTx gene sequences found in Elapidae. Three-finger toxins are usually composed of three exons and two introns; however, denmotoxin was found to have an additionally exon. Exon shuffling is a general mechanism for the creation of new genes (Pawlak and Kini 2008), and in the case of SVMPs, loss of exon segments has contributed to the evolution of the P-II and P-I classes.

Conclusions and Future Directions

The Promise of Venomics, Transcriptomics and Genomics

With the advancements in venom extraction methods for rear-fanged venomous snakes and in proteomic and molecular techniques, it is now possible to complete full analyses on rear-fanged snake venoms. This work will not only allow us to identify compounds that constitute an entire venom for a species, but will also provide a detailed description of venom composition of rear-fanged snakes that may be of medical significance with regards to snakebite. Further, rear-fanged venomics may provide insights into the utilization of these proteins for therapeutic drug development.

Proteomics, especially the high-throughput venomic and shotgun proteomic methods, have increased the sensitively and speed at which a whole venom can be characterized and the abundances of individual venom protein families determined. Although the vast majority of rear-fanged snakes may be considered as non-threatening to humans, proteomics allows identification of venom protein families which can further assist with examining the evolutionary relationships among venomous snakes and their toxins. For example, proteomic screening of the venom of *Thamnodynastes strigatus* indicated that the most abundant protein family consisted of a new kind of matrix metalloproteinase unrelated to the traditional SVMPs documented in all families of venomous snakes (Ching et al. 2012). This same study further identified the presence of a lactadherin-like factor V/VIII C-terminal domain as a part of the proteome, in

addition to well-known venom protein families such as SVMPs, C-type lectins, and CRiSPs. Rear-fanged snake venoms contain many of the same venom protein families that are found in medically significant venomous snakes, such as Viperidae and Elapidae, and as more transcript sequences are acquired and matched to corresponding purified proteins, in combination with biochemical and pharmacological assays, recurring venom protein domains can begin to be better defined and associated more directly with specific activities and molecular mechanisms of action.

Transcriptomics in combination with proteomics offers the ability to specifically identify the abundance of each venom protein, even between isoforms, from MS/MS peptide spectra and provides greater sequence coverage than what can be accomplished using only *de novo* sequence determination methods. Comparisons between proteomes and transcriptomes also provide insight into the translational regulation of venom proteins. This 'omic' approach was used to discover that non-coding RNA, including the important class of small posttranscriptional regulatory microRNAs, could potentially explain ontogenetic translational regulation. MicroRNAs produced in rear-fanged venom glands remain unexplored, though microRNAs have begun to be examined in Elapidae and Viperidae (Durban et al. 2013; Vonk et al. 2013). Transcriptomes assembled from a variety of snake body tissues as well as venom gland tissue of rear-fanged snake species will provide venom protein and non-venom homolog transcript sequences. These sequences can then be used to construct gene trees in order to reveal complete venom protein evolutionary histories.

Genomics in combination with transcriptomics allows for the exploration of transcriptional regulation seen within venom protein families and the unique evolution of venom protein genes. Genomics has revealed the dynamic evolution and adaptation of the venom system, such as the massive and rapid expansion of venom gene families that correlates with their functional importance in prey capture (Vonk et al. 2013). The expansion of venom protein gene families may occur in response to an evolutionary arms race between venomous snakes and their prey. Venom gene sequences are therefore ideal for studying accelerated patterns of evolution and the association between genotype and adaptive phenotypes.

The combination of proteomics, transcriptomics, and genomics in the study of rear-fanged snake venoms can provide a holistic approach to understanding venom protein evolution and regulation, which in turn impacts overall crude venom composition. However, it is also important to view the evolution and expression of rear-fanged venom proteins in terms of biological significance. Venom is a trophic adaptation and as such, the presence and abundance of these proteins relates directly to prey capture and organism natural history.

Forgotten Aspects of Understanding Venom Evolution

Snake venom prey-specific toxins provide an ideal model to study genotype and phenotypic fitness interactions because the function of phenotypic variation (venom composition) can be related to the nature of the adaptation

(prey preference and susceptibility). At the molecular level, there is the evolution of protein catalytic/ligand-binding sites and targeting. At the organismal level, there are selection pressures brought on by prey availability, preference and susceptibility to specific toxin effects. Therefore, the biological roles of venom proteins should be incorporated into high-throughput proteomic, transcriptomic, and genomic results aimed at understanding venom evolution.

Rear-fanged venomous snakes encompass several families and subfamilies of the Colubroidea, and collectively they include the largest number and diversity of venomous snakes. To explore the biological roles of rear-fanged snake venoms, or individual toxins within these venoms, it is important to use adequate toxicity models for assays that match the biology of the snake being studied. Understanding the diversity of venom components and their differential effects towards specific prey will facilitate a greater understanding of the selective mechanisms driving snake venom evolution and adaptation (Mackessy et al. 2006). Future studies should take into account the interactions between the snake's venom and its natural prey, since toxicity is best defined within the context that it is being used. There is a need for inbred non-mammalian vertebrate species to be used as models for LD_{50} assays, as well as for viable non-vertebrate models of whole organism toxicity. Such models would be ideal for toxinologists interested in receptor- ligand evolution and positive selection of venom proteins involved in coevolutionary predator/prey arm races (Mackessy 2002). Some of these toxins may prove useful for understanding diversification and evolution of important ion channels, such as the nicotinic acetylcholine

receptor; based on selective toxicity of *B. irregularis* venom and irditoxin, there appears to have been some form of coevolutionary adjustment between predator armaments (venom) and prey susceptibilities (receptor binding) which imparts particular effectiveness against specific prey types (Mackessy et al. 2006; Pawlak et al. 2009). The differential activities of these and other toxins could be exploited for the development as molecular tools for dissecting receptor-ligand binding interactions, and they may provide clues for their exploitation as therapeutics.

Potential for Drug Development from Venom Proteins and Peptides

A number of toxins have proved to be excellent research tools to decipher the molecular details of physiological processes, and several have led to the development of novel therapeutic agents (Lewis 2009; Takacs and Nathan 2014). Captopril, the first successful venom-based drug, was developed from bradykinin-potentiating peptides from the venom of *Bothrops jararaca* (Lancehead Viper) and is still on the market as an anti-hypertensive drug. Other venom-based drugs include tirofiban (aggrastat) and integrilin (eptifabatide) that were both designed from the structure of snake venom disintegrins (Vonk et al. 2011; Saviola et al. 2014). There are many rear-fanged venomous snake species that have venoms yet to be studied, providing an untapped source of proteins with novel activities for therapeutic development (Saviola et al. 2014).

Recently, five venoms from rear-fanged snake species were evaluated for potential anti-leishmanial activity. Exposure to relatively high levels of these rear-fanged snake venoms resulted in cytotoxicity towards cultured promastigote states of *Leishmania major* and venom of one species, *T. b. lambda*, showed significant cytotoxicity even at lower doses (Peichoto et al. 2011a). Because rear-fanged snake venoms contain many of the same venom protein families as front-fanged venomous snakes, and because structural motifs of venom proteins are conserved but possess activities and specificities that may be highly variable, exploration of rear-fanged snake venom proteins could uncover some highly useful compounds. Anti-coagulants in rear-fanged snake venoms include SVMPs, serine proteases, and phospholipase A_2 enzymes (Saviola et al. 2014), and initial analyses indicate that at least some may show higher specificities than homologs from front-fanged snake venoms (Weldon and Mackessy 2012). These variants provide opportunities to decipher the subtleties in functional sites in order to understand the plasticity of venom protein structure and function. Venom proteins can serve as templates for biomedical engineering and provide insight into selective receptor binding (Kini and Doley 2010), as is exhibited by several 3FTxs from rear-fanged snake venoms. Without the selectivity of alpha-bungarotoxin, a 3FTx from venom of *Bungarus multicinctus* (Many-banded Krait), the knowledge of the distribution of nicotinic acetylcholine receptors and neurotransmitter communication wouldn't have advanced to its current state. Because venom proteins originated from ancestral proteins that served cellular physiological (housekeeping) roles and often have exceptionally high binding

specificities, they are ideal candidates for unraveling cellular signaling pathways and the resulting disruption of these pathways. Rear-fanged snake venoms provide a largely unexplored resource of toxins, and with recent advances in proteomic, transcriptomic, and genomic approaches, there should be an increase in future research focused on these venoms.

CHAPTER II

AN ANALYSIS OF VENOM ONTOGENY AND PREY-SPECIFIC TOXICITY IN THE MONOCLED COBRA (*NAJA KAOUTHIA*)

Cassandra M. Modahl[a], Ashis K. Mukherjee[ab], and Stephen P. Mackessy[a]
[a]School of Biological Sciences, University of Northern Colorado, 501 20th St., Greeley, CO 80639-0017, USA
[b]Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, India

Abstract

Venoms of snakes of the family Elapidae (cobras, kraits, mambas, and relatives) are predominantly composed of numerous phospholipases A_2 (PLA_2s) and three-finger toxins (3FTxs), some of which are lethal while others are not significantly toxic. Currently, the only identified prey-specific toxins are several nonconventional 3FTxs, and given the large diversity of 3FTxs within *Naja kaouthia* (Monocled Cobra) venom, it was hypothesized that several 3FTxs, previously found to be non-toxic or weakly toxic 3FTxs in murine models, could potentially be toxic towards non-murine prey. Additionally, it was hypothesized that ontogenetic dietary shifts will be correlated with observable changes in specific 3FTx isoform abundance. Adult and juvenile *N. kaouthia* venom composition was investigated using ion-exchange FPLC, 1D and 2D SDS-PAGE, mass spectrometry, and various enzymatic and LD_{50} assays. Alpha-cobratoxin (α-elapitoxin) was determined to be the only significantly toxic (LD_{50} < 1 µg/g)

3FTx found in *N. kaouthia* venom and was equally toxic toward both lizard and mouse models. The abundance and diversity of 3FTxs and most enzyme activities did not vary between adult and juvenile cobra venoms; however, total venom PLA_2 activity and specific PLA_2 isoforms did vary, with juveniles lacking several of the least acidic PLA_2s, and these differences could have both biological (related to predation) and clinical (antivenom efficacy) implications. However, the ubiquitous presence of α-cobratoxin in both adult and juvenile cobra venoms, showing high toxicity toward both reptile and mammal models, represents a venom compositional strategy wherein a single potent toxin effectively immobilizes a variety of prey types encountered across life history stages.

Introduction

One of the most abundant cobra species, *Naja kaouthia* (Monocled Cobra), has a range that includes India, Bangladesh, Nepal, Myanmar, southwestern China, and Thailand (Mukherjee and Maity 2002; Reali et al. 2003). In Thailand, snakebite envenomations by *N. kaouthia* account for the highest number of human fatalities among all venomous snake species (Kulkeaw et al. 2007). Patients who have systemic envenoming by *N. kaouthia* usually develop neurotoxic symptoms, including ptosis, dysphagia, and increased salivation, followed by coma and death from respiratory paralysis in severe cases (Sells et al. 1994; Reali et al. 2003).

Because *N. kaouthia* is very common and is responsible for human morbidity and mortality, there have been many studies published characterizing

specific venom proteins and describing overall venom composition (Hamako et al. 1998; Sakurai et al. 2001; Meng et al. 2002; Mukherjee and Maity 2002; Doley and Mukherjee 2003; Kulkeaw et al. 2007; Mordvintsev et al. 2007; Mordvintsev et al. 2009; Debnath et al. 2010), including two recent publications on the complete venomics profile of *N. kaouthia* (Laustsen et al. 2015; Tan et al. 2015). The venom of *N. kaouthia* is primarily composed of three-finger toxins (3FTxs), with either neurotoxic or cardiotoxic/cytotoxic activities, and phospholipase A_2 (PLA_2) isoforms (Namiranian and Hider 1992; Kulkeaw et al. 2007; Laustsen et al. 2015; Tan et al. 2015). Geographic venom variation was also recently documented for *N. kaouthia* (Tan et al. 2015), but because pooled venoms were used in these studies, there is still a lack of information on individual intraspecific venom variation or ontogenetic venom compositional changes in *N. kaouthia*.

Venom variability has been documented at the family, genus, species and intraspecific levels (Chippaux et al. 1991; Mackessy 2010b). Intraspecific venom variation may occur between individuals of different geographic locations, dietary habits, genders and age (Minton and Weinstein 1986; Mackessy 1988; Chippaux et al. 1991; Mackessy 1993a; Daltry et al. 1996a; Daltry et al. 1996b; Mackessy et al. 2003; Menezes et al. 2006; Alape-Giron et al. 2008). Ontogenetic venom variation studies have largely focused on the subfamily Crotalinae (pit vipers), *Crotalus* and *Bothrops* species in particular (Meier 1986; Minton and Weinstein 1986; Mackessy 1988; Gutiérrez et al. 1991; Mackessy 1993a; Saldarriaga et al. 2003; Alape-Giron et al. 2008; Zelanis et al. 2009). A few studies have analyzed

ontogenetic venom variation in the Elapidae (cobras, kraits, mambas, and relatives), but the conclusions of these studies varied (see below).

Several studies of Australian elapids found that there was no significant difference in the venom composition of juvenile and adult *Oxyuranus scutellatus* (Coastal Taipans), *Oxyuranus microlepidotus* (Inland Taipans), and *Notechis scutatus* (Tiger Snakes) (Tan et al. 1992; Tan et al. 1993b, a). Minton (1967) found that venom toxicity, hemorrhagic activity, indirect hemolysin, direct hemolysin, and hemagglutinin for *Naja naja ssp.* (Common Cobra) increased with age (Minton 1967). Unfortunately, since this study was conducted before the more recent revisions of the genus *Naja* that revealed at least ten species of *Naja* throughout India and southeast Asia (Wüster 1996), it is unknown which species was actually studied. Meier and Freyvogel (1980) found a decrease in venom toxicity with age for *N. nigricollis* (Black-necked Spitting Cobra) (Meier and Freyvogel 1980). A more recent study on ontogenetic venom variation in *N. atra* (Chinese Cobra) found higher phosphomonoesterase and L-amino acid oxidase activity and lower nucleotidase, PLA_2, hyaluronidase, and fibrinolytic activity within neonate *N. atra* venom (He et al. 2014).

Ontogenetic changes in venom composition have been found to correlate with snake dietary shifts (Mackessy 1988; Andrade and Abe 1999; Mackessy et al. 2006; Zelanis et al. 2009). In *Crotalus oreganus helleri* (Pacific Rattlesnakes) (Mackessy 1988) and *Bothrops jararaca* (Jararacas), an ontogenetic shift in diet from ectothermic prey (arthropods, lizards, and amphibians) as a juvenile to endothermic prey (mammals) as an adult was associated with a shift in venom

toxicity, with juvenile venom being more toxic to ectotherms and adult venom being more toxic towards mammals (Andrade and Abe 1999). A similar pattern was noted for *Boiga irregularis*, with juvenile venoms more toxic than adult venoms towards geckos, suggestive that this trend is not just found in vipers (Mackessy et al. 2006).

In their first year, *N. kaouthia* feed primarily on frogs and newborn rats, adult snakes feed on adult rats, snakes, lizards, fish, birds, and bird eggs (Chaitae 2000). Therefore, as *N. kaouthia* ages, a wider diversity of prey is taken, most likely because juveniles are gape-limited to feed on smaller prey. Prey-specific toxins have not been identified in cobras, even though 3FTxs that are weakly toxic towards murine models have been identified in cobra venoms, and 3FTxs are currently the only venom proteins to be directly linked to prey-specific toxicity (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013). Non-conventional/weak 3FTxs exist in *N. kaouthia* venom, and this subclass of 3FTxs have been suggested to confer prey-specific toxicity in *Ophiophagus hannah* venoms (Chang et al. 2013), but this hypothesis has not been tested.

Documentation of intraspecific variation in venoms also has important clinical and antiserum production applications. Intraspecific venom variation in the *Naja naja* (Spectacled Cobra) has resulted in the manifestation of different clinical envenomation symptoms and a lack of antivenom efficacy between India and Sri Lanka *N. naja* populations (Kularatne; et al. 2009). The present study compares the biochemical composition of adult and juvenile *N. kaouthia* venoms and addresses the following questions: Are there differences in venom protein

content and activity between adult and juvenile cobras? Is there a difference in the toxicity of adult and juvenile cobra venoms towards ectothermic (lizards) and endothermic (mammalian) prey? Are there components of *N. kaouthia* venom, such as nonconventional 3FTxs, that exhibit prey-specific toxicity? It was hypothesized that non-conventional 3FTxs in *N. kaouthia* venoms would exhibit selective toxicity toward non-mammalian prey.

Materials and Methods

Reagents

Reagents for protein concentration assays were purchased from BioRad Inc. (San Diego, CA, U.S.A). Precast NuPAGE 12% Bis-Tris mini gels, Novex Mark 12 unstained molecular mass standards, LDS sample buffer and MES running buffer were purchased from Life Technologies (Grand Island, NY, U.S.A). Two-dimensional gel electrophoresis supplies, including DeStreak rehydration solution, IPG pH 3-11 buffer and Immobiline DryStrip pH 3-11, were purchased from GE Healthcare (Pittsburgh, PA, U.S.A). Phospholipase A_2 assay kit was purchased from Cayman Chemical Co. (Ann Arbor, MI, U.S.A). Azocasein, DTT, DTNB, acetylthiocholine iodide, bis-p-nitrophenylphosphate, L-kynurenine, human fibrinogen and all other reagents (analytical grade or better) were obtained from Sigma-Aldrich (St. Louis, MO, U.S.A).

Venoms and Animals

Pooled and lyophilized adult (over three years in age) and 10 individual juvenile (less than three months in age) *N. kaouthia* venoms were donated by the Kentucky Reptile Zoo (Slade, KY, U.S.A). All *N. kaouthia* were of Thailand origin

and kept under the same husbandry conditions. The pooled adult venom included parents of the juveniles to reduce the potential effects of venom variation due to genetic variability. Venoms were stored frozen at -20 °C until needed. Venom protein concentration was determined using bovine gamma globulin standard and the method of Bradford (1976) as modified by BioRad Inc. (Bradford 1976). The calculated protein concentrations for each crude venom sample were used in all other analyses and enzymatic activity calculations. NSA mice (*Mus musculus*) were bred in the University of Northern Colorado (UNC) Animal Resource Facility and House Geckos (*Hemidactylus frenatus*) were obtained from Bushmaster Reptiles (Longmont, CO, USA); all procedures were reviewed and approved by the UNC IACUC (protocol 1504D-SM-SMLBirds-18).

One-Dimensional Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE on NuPage 12% Bis-Tris mini gels (Life Technologies, Inc., U.S.A) was performed to compare the relative molecular masses of venom components of adult and juvenile *N. kaouthia*. Samples and buffers were prepared under reducing conditions according to the manufacturer. Venom samples were run at 10 µg and 20 µg per lane, with 5 µL of Novex Mark 12 unstained mass standard in one lane for estimation of molecular masses. The gel was run at 180 V, stained with 0.1% Coomassie Brilliant Blue R250 overnight, destained (50/40/10, v/v, ddH_2O:methanol:glacial acetic acid) for two hours, and imaged.

Two-Dimensional Polyacrylamide Gel Electrophoresis (2D SDS-PAGE)

Two-dimensional gel electrophoresis was performed to evaluate protein composition differences between adult and juvenile *N. kaouthia* venoms. Venom (170 µg) was added to 135 µL DeStreak rehydration solution containing 18 mM DTT and 0.5% IPG pH 3-11 buffer (GE Healthcare, Pittsburgh, PA, U.S.A). This solution was added to the ceramic holder, overlain with a 7 cm IPG pH 3-11 strip and then mineral oil, and placed on an IPGphor Electrofocusing System (GE Healthcare, Pittsburgh, PA, U.S.A). The following isoelectric focusing steps were performed: rehydration for 12 hours, 500 V for 30 minutes, 1000 V for 30 minutes, 5000 V for 3 hours, and a 40 V hold. For the second dimension electrophoresis, the focused IPG strip was equilibrated in a reduction buffer (I mL 4X LDS buffer, 400 µL 150 mM DTT [final concentration 15 mM], and 2.6 mL ddH_2o) at ambient temperature for 10 minutes and then equilibrated in an alkylation buffer (1 mL 4X LDS, 3 mL ddH_2O, and a final concentration of 60 mM iodoacetamide) at ambient temperature for 10 minutes. SDS-PAGE was carried out using NuPage 12% Bis-Tris 2D well gels following the same procedure as described above for 1D SDS-PAGE.

Enzyme Assays

Metalloproteinase activity of adult and juvenile crude venoms (80 µg) were determined using azocasein as a substrate (Aird and da Silva 1991), and activity was expressed as $\Delta A_{342\ nm}$/min/mg venom protein. L-amino acid oxidase activity was assayed according to Weissbach et al. (1961) with 80 µg of venom, and the

activity was expressed as nmol product formed/min/mg protein (Weissbach et al. 1960). Phospholipase A_2 activity was determined using a commercially available kit (Cayman Chemical Co., Michigan, USA) as described by the manufacturer using 2 µg venom in 200 µL total volume. Absorbance was measured at 414 nm every minute for five minutes at 37ºC and the linear portion of the curve was used to express activity as µmol product formed/min/mg protein. Phosphodiesterase activity was assayed with 80 µg venom using 1.0 mM bis-p-nitrophenylphosphate as a substrate, following the protocol of Laskowski (1980); activity was reported as $\Delta A_{400\ nm}$/min/mg protein (Laskowski 1980). Acetylcholinesterase activity was determined using venom (5 µg) incubated with the 7.5 mM acetylthiocholine iodide substrate (21.67 mg/ml) in a cuvette at 37°C (Ellman et al. 1961). Absorbance at 412 nm was recorded every 10 seconds for five minutes. The linear portion of the graph produced was used to calculate specific activity as µmol of thiocholine produced/min/mg venom protein. Fibrinogenase activity was determined using 20 µg of venom incubated with human fibrinogen (final concentration 0.5 mg/mL in 100 mM Tris-HCl, pH 8.0) in a total volume of 200 µL for periods of 0, 1, 5, 10, 30 and 60 minutes (Ouyang and Huang 1979). Twenty µL of this reaction mixture was removed at each time point, mixed with an equal volume of 4% SDS and 5% 2-mercaptoethanol, and then heated in boiling water for 10 minutes. Five µL aliquots were combined with 2x LDS buffer, electrophoresed on a 12% NuPAGE Bis-Tris gel and processed as above. All assays were performed in duplicate for the pooled adult *N. kaouthia* venom and for each individual juvenile venom sample.

Cation-Exchange Fast Protein Liquid Chromatography (FPLC)

Cation-exchange FPLC was also used to compare venom compositional differences between adult and juvenile *N. kaouthia* venoms. Crude venom (3 mg) in a volume of 500 µl of buffer (20 mM MES, pH 6.0) was injected onto a GE Healthcare Tricorn Mono S 5/50 GL column (5 x 50 mm) equilibrated with 20 mM MES, pH 6.0. The elution was performed at a flow rate of 1 mL/min at ambient temperature with a linear NaCl gradient from 0 to 0.4 M over 60 minutes and then 1M NaCl for 5 minutes. Eluted proteins were monitored at 280 nm. The protein composition of each peak was evaluated by SDS-PAGE and MALDI-TOF MS.

Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight Mass Spectrometry (MALDI-TOF)

MALDI-TOF MS was used to determine the molecular masses of all venom proteins found in the FPLC peaks. Fractions from all major Mono S FPLC peaks were analyzed with a Bruker Ultraflex MALDI-TOF mass spectrometer (Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, CO, U.S.A) operating in linear mode. Approximately 1 µg of protein (1.0 µL) from each peak was spotted onto a sinapinic acid matrix (10 mg/mL in 50% acetonitrile, 0.1% trifluoroacetic acid (TFA); 1.0 µL), an isopropanol wash (2.5 µL) was used to remove salts, and the spectra were acquired in the mass range of 3.0-25 kDa. These protein masses were then matched to previously identified *N. kaouthia* venom proteins (Kulkeaw et al. 2007; Laustsen et al. 2015).

Reverse-Phase High Performance Liquid Chromatography (RP-HPLC)

A Waters RP-HPLC (Empower software) was used to purify proteins from adult *N. kaouthia* venom that were used in toxicity assays. Fractions (2 mL total) from all Mono S FPLC major peaks were injected onto a Phenomenex Jupiter 5 µm C18 (250 x 4.60 mm) column connected to Waters 515 HPLC pumps. The column was equilibrated with 0.1% TFA (buffer A) at a flow rate of 1 mL/min. Elution was carried out with a linear gradient of 0-35% buffer B (80% acetonitrile containing 0.01% TFA) over 5 minutes, then a linear gradient of 35%-60% buffer B over 30 minutes, and 100% buffer B for the remaining 10 minutes at room temperature (~23 ◦C). The flow rate was 1 mL/min and protein elution was continuously monitored at 280 and 220 nm with a Waters 2487 Dual λ absorbance detector. Fractions associated with the most abundant protein observed in each chromatogram were lyophilized, assayed for protein concentration (Bradford,1976, as modified by BioRad Inc.) and purity evaluated by SDS-PAGE analysis.

Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

To determine the identities of toxins used for LD_{50} assays, each RP-HPLC purified protein peak was analyzed by LC-MS/MS, performed at the Florida State University College of Medicine Translational Science Laboratory (Tallahassee, FL, U.S.A). Samples were digested using the Calbiochem ProteoExtract All-in-

one Trypsin Digestion kit (Merck, Darmstadt, Germany) with LC/MS grade solvents according to the manufacturer's instructions. The LC-MS/MS analyses were performed using an LTQ Orbitrap Velos equipped with a Nanospray Flex ion source and interfaced to an Easy nanoLC II HPLC (Thermo Scientific). Peptide fragments were separated using a vented column configuration consisting of a 0.1 x 20 mm, 3 µm C18 trap column and a 0.075 x 100 mm, 3 µm C18 analytical column (SC001 and SC200 Easy Column respectively, Thermo Scientific). The elution gradient consisted of 5% buffer B (0.1% formic acid in HPLC grade acetonitrile) and 95% buffer A (0.1% formic acid) at the run start, to 35% B at 60 min, to 98% B from 63-78 min with a flow rate of 600 nl/min from 64-78 min, and 5% B at 300 nl/min at 79 min. The mass spectrometer was operated in positive mode nanoelectrospray with a spray voltage of +2300 V. A "Top 9" method was used with precursor ion scans in the Orbitrap at 60K resolving power and fragment ion scans in the linear ion trap. Precursor ion selection using MIPS was enabled for charge states of 2+, 3+ and 4+. Dynamic exclusion was applied for 60 sec at 10 ppm. ITMS scans were performed using collision-induced dissociation (CID) at 35% normalized collision energy. MS/MS peptide spectra produced were interpreted using Mascot (Matrix Science, London, UK; version 1.4.0.288), Sequest (Thermo Fisher Scientific, San Jose, CA, U.S.A; version 1.4.0.288), and X! Tandem (thegpm.org; version CYCLONE 2010.12.01.1), assuming a trypsin digestion. The Mascot5_Trembl_bony vertebrate database, and the Sequest and X! Tandem Uniprot Serpentes (A8570) databases, were used for homology searches. Sequest and X! Tandem were searched with a

fragment ion mass tolerance set to 0.6 Da and a parent ion tolerance of 10 ppm. Mascot was searched with a fragment ion mass tolerance of 0.8 Da and a parent ion tolerance of 10 ppm. Glu→pyro-Glu of the N-terminus, ammonia loss of the N-terminus, Gln→pyro-Glu of the N-terminus, carbamidomethylation of cysteines and carboxymethylation of cysteines were specified as variable post-translational modifications within X! Tandem. Oxidations of methionine, carbamidomethyl cysteine, and carboxymethyl cysteine were specified as variable post-translational modifications within Mascot and Sequest. Results were viewed and validated within Scaffold (Proteome Software Inc., Portland, OR, U.S.A; version 4.4.6), and protein identities were accepted if they could be established at >99.9% probability and contained at least one identified peptide. Given the sensitivity of the MS/MS instrument, purified protein identities were also restricted to those proteins that consisted of the largest number of fraction spectra, after normalization to protein molecular mass as determined by MALDI-TOF MS.

Lethal Toxicity (LD_{50}) Assays

Comparative toxicity of adult and juvenile *N. kaouthia* venoms was evaluated (LD_{50} assays) using a non-model species, House Gecko (*Hemidactylus frenatus*), and a standard model species, NSA mouse (*Mus musculus*). In addition, LD_{50} assays were also performed for RP-HPLC purified venom proteins from adult *N. kaouthia* venom to determine if any toxin exhibited prey-specific toxicity. Doses used for crude venoms were 0.2, 0.6, 0.8, 1.0, 1.2 and 1.5 µg/g (in 50 µL) for lizards and 0.2, 0.4, 0.5, 0.6, 0.8 and 1.0 µg/g (in 100 µL) for mice. Doses for individual purified proteins were 0.1, 0.5, 1, and 5 µg/g in

lizards; toxicities toward mice were taken from the literature (Karlsson 1973; Fryklund and Eaker 1975; Joubert and Taljaard 1980b, c; Joubert and Taljaard 1980a; Laustsen et al. 2015). All doses were injected intraperitoneally in sterile PBS, and three animals per dose were used; three mice and three lizards were injected with only PBS as controls. Doses were adjusted to individual animal body masses, with lethality expressed as micrograms of venom per gram body mass (μg/g) producing 50% mortality after 24 hours (Reed and Muench 1938). All procedures were reviewed and approved by the UNC-IACUC (protocol 1504D-SM-SMLBirds-18).

Data Analysis

A series of single-sample t-tests were performed in R (version 3.1.1) for each crude venom enzyme activity to determine if there was a significant difference between adult and juvenile *N. kaouthia* venoms. Single-sample t-tests were chosen instead of independent t-tests because there was only one pooled sample for the adult *N. kaouthia* venom (n=1). Therefore, the adult sample was used as the reference mean for the comparison between the adult and juvenile samples. For the LD_{50} assays, the nonparametric Spearman-Karber method was used for both LD_{50} value estimations and determination of 95% confidence intervals (Finney 1978). If confidence intervals overlapped, it was determined that there was not a significant difference between the values being compared.

Results and Discussion

One-Dimensional Gel Electrophoresis (SDS-PAGE)

There were several qualitative differences in protein composition noted between adult and juvenile *N. kaouthia* reduced SDS-PAGE venom profiles. Approximately 18 protein bands between 4-100 kDa were observed in *N. kaouthia* venoms (Fig. 1). Bands around 31-36.5 kDa were present in several juvenile venoms but were completely absent in the pooled adult *N. kaouthia* venom. There were also two bands in juvenile venoms around 25 kDa (typically cysteine-rich secretory proteins), and only one was apparent in the adult venoms. A striking difference between adult and juvenile venoms was the low abundance in juvenile venoms of bands around 13 kDa (typically PLA_2s). These data indicate that juvenile venoms differ in several respects from adult *N. kaouthia* venoms.

Recently, Tan et al. (2015) completed proteomes for venoms of *N. kaouthia* from three different localities using a venomics methodology (Calvete 2013). Venoms from *N. kaouthia* originating from Thailand were included in the Tan et al. (2015) study; therefore, this publication was used as an additional reference for the identification of reduced SDS-PAGE protein bands for the present study (Fig. 1). The bands at 31-36.5 kDa observed in some juvenile venoms and absent from the adult venom are potentially cobra venom factor γ-chains (Kulkeaw et al. 2007; Tan et al. 2015). Cysteine-rich secretory proteins (CRiSPs) were present at a mass of approx. 25 kDa for Thailand *N. kaouthia*

(Tan et al. 2015); these proteins frequently occur at this mass range in many reptile venoms (Mackessy 2010b). The two bands in this range for juveniles indicate that some juvenile venoms have two abundant CRiSP isoforms, whereas the adult venoms have only one. The low abundance in juvenile venoms of bands at approx.13 kDa are in the mass range expected for PLA_2s (Mackessy 2010b; Tan et al. 2015), indicating a lower abundance to near complete absence of PLA_2 enzymes in some juvenile *N. kaouthia* venoms (Fig. 1).

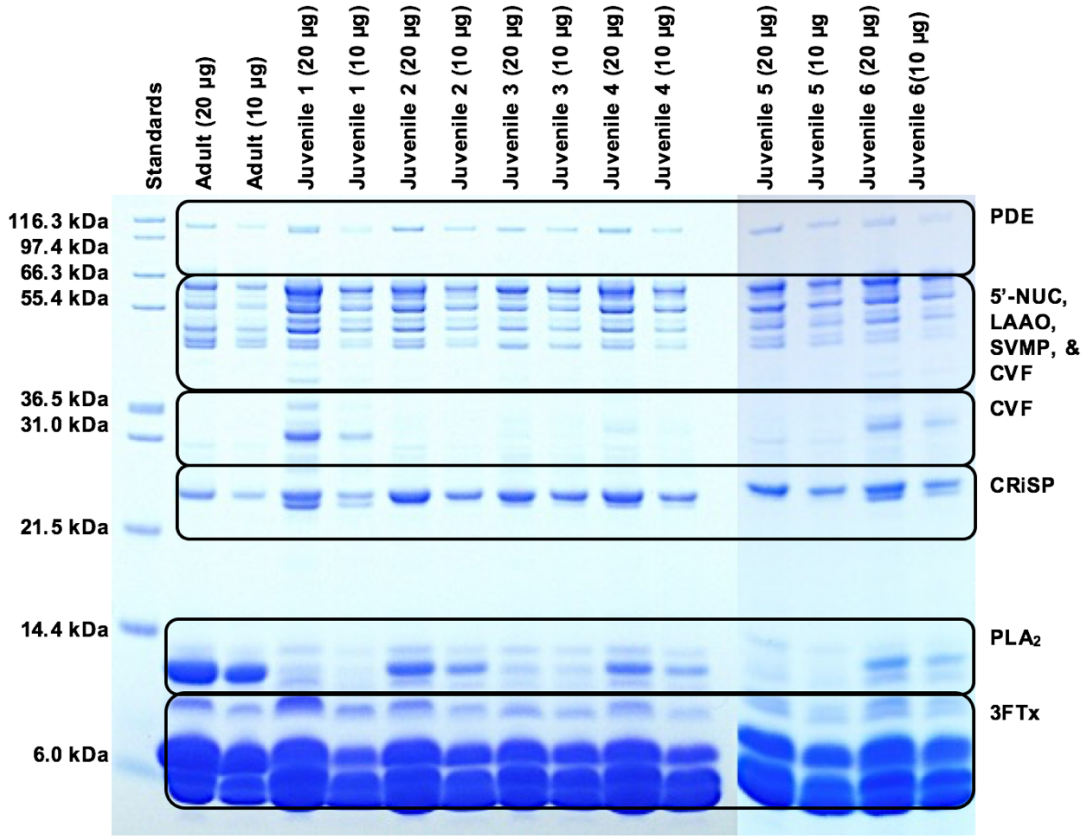


Figure 1. SDS-PAGE comparison of *N. kaouthia* reduced venoms from adult (pooled) and individual juvenile snakes. Molecular mass standards are indicated in the far left lane. Protein families are identified on the far right and include PDE (phosphodiesterase), 5'-NUC (5'-nucleotidase), LAAO (L-amino acid oxidase), SVMP (snake venom metalloproteinase), CVF (cobra venom factor), CRiSP (cysteine-rich secretory protein), PLA_2 (phospholipases A_2) and 3FTx (three-finger toxin). Identities are based on previously published work (Mackessy 2010b; Tan et al. 2015).

Two-Dimensional Polyacrylamide Gel Electrophoresis (2D SDS-PAGE)

A 2D gel electrophoretic profile of pooled adult *N. kaouthia* venom is shown in comparison to one juvenile venom (Fig. 2). There were five major differences between the gel profiles. First, there is a larger abundance of high molecular mass proteins (approximately around 67 kDa) within the juvenile venom. There are also proteins present (approx. 36.5 kDa) in the juveniles that are absent in the adults, and two proteins (at approx. 14 kDa and 6 kDa) are absent from the juvenile venom but present in abundance in the adult *N. kaouthia* venom. Lastly, there are differences in abundances of basic low molecular mass proteins (most likely 3FTxs) clustered on the bottom left (basic) edge of the gels.

The proteome of Thailand *N. kaouthia* venom was described previously using two-dimensional gel electrophoresis to separate all venom proteins by molecular mass and isoelectric point, and proteins were then identified using tandem mass spectrometry of tryptic peptides (Kulkeaw et al. 2007). Kulkeaw et al. (2007) found that cobra venom factor contains three subunits (α-, β-, γ-chains) that will appear on a 2D SDS-PAGE gel at apparent molecular masses of approximately 70, 50 and 30 kDa. Using Kulkeaw et al. (2007) as a reference, the larger mass proteins that are most abundant in the juvenile venoms are likely cobra venom factor (Fig. 2). This supposition is also corroborated by results observed with 1D SDS-PAGE gels in the present study.

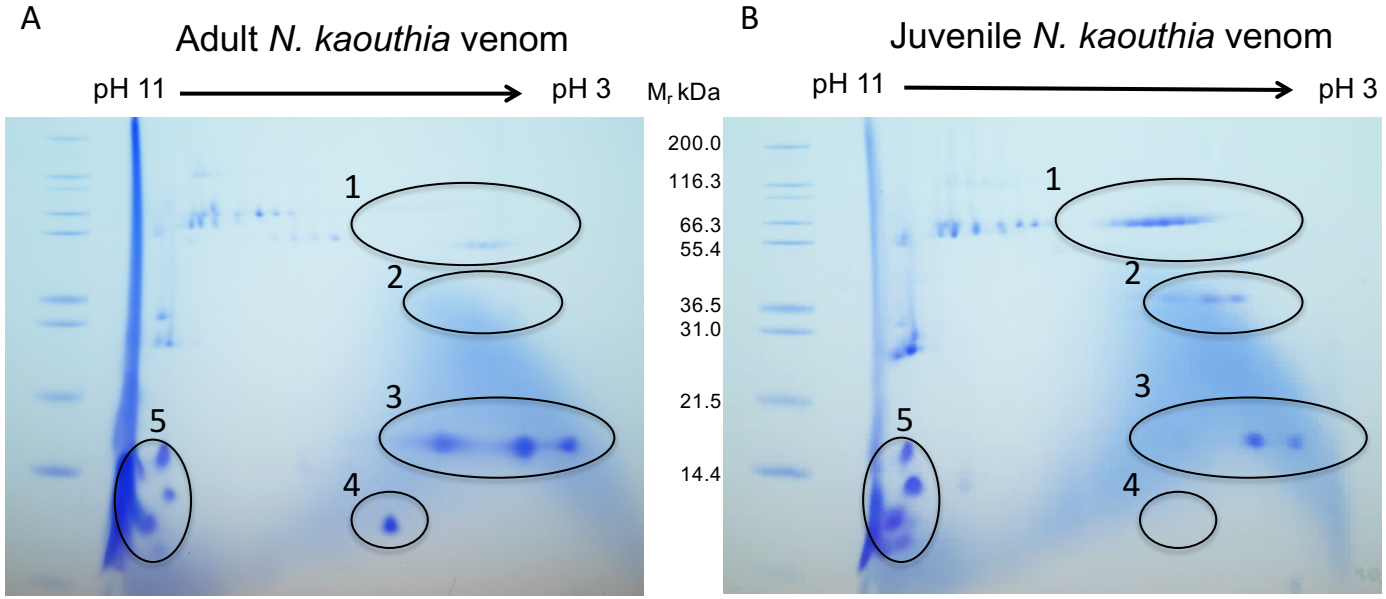


Figure 2. Comparison of 2D gel electrophoresis profiles of adult and juvenile *Naja kaouthia* venoms. Venoms (pooled; 170 µg) from adult (A) and from one juvenile (NK-J1; 170 µg) (B) N. kaouthia were profiled under the same conditions (pH 3-11, 12% gel); these samples were also subjected to FPLC/HPLC fractionation (see below). Circled areas highlight differences in abundance of 1) cobra venom factors, 2) unidentified proteins 3) PLA2s, 4) a potential 3FTx, and 5) a cluster of 3FTxs (Kulkeaw et al. 2007). The juvenile venom profile shown highlights the more prominent PLA_2 differences.

Proteins of approx. 36.5 kDa were not identified by Kulkeaw et al. (2007); these proteins occur in juvenile *N. kaouthia* venom but are lacking in adult venom. Proteins with masses of approx. 13-14 kDa were identified as PLA_2s (Kulkeaw et al. 2007), and it is evident that juvenile *N. kaouthia* venom lacks the distinct PLA_2 cluster (specifically the most basic isoform) that is abundant in adult *N. kaouthia* venoms. In general, there is also a lower abundance of PLA_2s within the juvenile venom profile, corresponding with differences in adult and juvenile venoms observed on 1D SDS-PAGE. The last difference is the presence of a lower molecular mass, somewhat acidic protein (6-7 kDa) in adult venom and absent from juvenile venom; differences in abundances of basic proteins within this mass range, most likely 3FTxs (Tan et al. 2015), also exist between juvenile and adult venoms.

Enzyme Assays

L-amino acid oxidase and metalloproteinase activities were not significantly different between adult and juvenile venoms ($p = 0.18$ and 0.14, respectively; Table 1). Metalloproteinase activity has previously been reported to be very low within elapid venoms in comparison to vipers, and the metalloproteinase activity reported in this study (using azocasein substrate) is similar to previously reported lower values using casein substrates (Aird and da Silva 1991; Das et al. 2013). Metalloproteinases in *N. kaouthia* venom have also been found to cleave platelet von Wilebrand factor (Hamako et al. 1998; Wijeyewickrema et al. 2007).

Table 1. Enzyme activity of adult and juvenile *Naja kaouthia* venoms.

	Adult *Naja kaouthia* venom	Juvenile *Naja kaouthia* venom†
Metalloproteinase activity (ΔA_{342nm}/min/mg)	0.004	0.005 ± 0.002
L-amino acid oxidase activity (nmol product/min/mg)	4.395	5.621 ± 2.668
Phospholipase A_2 activity (µmol product/min/mg)	**2.216***	**0.643* ± 0.742**
Phosphodiesterase activity (ΔA_{400nm}/min/mg)	**0.337***	**0.566* ± 0.123**
Acetylcholinesterase activity (µmol product/min/mg)	**7.202***	**4.682* ± 2.485**

* Significantly different by single sample t-test ($p < 0.05$); †, n = 10

Phospholipase A_2, phosphodiesterase and acetylcholinesterase activities were significantly different between adult and juvenile venoms ($p = 0.00005$, 0.0002 and 0.01, respectively; Table 1). Consistent with the results observed in this study, significantly lower acetylcholinesterase and phospholipase A_2 activity has been reported in neonate *Naja atra* (Chinese Cobras) when compared to adults (He et al. 2014). PLA_2 activity in adult *N. kaouthia* originating from India (Das et al., 2013) was slightly higher than the value reported in the present study; however, PLA_2 activity appears to be quite variable among cobra venoms (Tan and Tan 1988), and the relative abundance of PLA_2 enzymes within *N. kaouthia* venom proteomes from different localities varied from 12.2% to 23.5% (Tan et al., 2015). Within these same *N. kaouthia* venom proteomes, phosphodiesterases accounted for only 0.3-0.4% of *N. kaouthia* venom, and acetylcholinesterases

were not present or were at levels below detection (Tan et al. 2015). However, some low abundance enzymes are difficult to detect using proteomic methodologies, but are detected using enzyme assays. It is possible that higher acetylcholinesterase activity within adult *N. kaouthia* venom could contribute to differential clinical symptoms.

PLA_2 activity, an important contributor to morbidities following envenomations by *N. kaouthia*, showed significant age-related differences (Table 1). The significantly lower PLA_2 activity of juvenile venoms also corresponds with the lower abundance and lack of PLA_2s observed in 1D and 2D SDS-PAGE juvenile venom profiles. Phospholipases A_2 in venoms often produce severe tissue damage as a result of cell membrane hydrolysis and release of autopharmacologically-active compounds such as arachidonic acid (Kini 2003; Doley et al. 2004; Mukherjee 2007). Results indicate that envenomations by juvenile *N. kaouthia* should result in significantly lower incidence of severe tissue damage compared to envenomations by adult cobras, which have a higher abundance of PLA_2 and therefore higher necrotizing potential.

Many snake venoms contain fibrinogenolytic enzymes, commonly serine proteinases, which superficially mimic the actions of thrombin on circulating fibrinogen (Mackessy 1993a, b, 2010c). Fibrinogen, composed of three subunits ($A\alpha$, $B\beta$, and γ), is typically converted by activated thrombin into fibrin during the formation of a blood clot via the selective cleavage of fibrinopeptides A and B from the α and β subunits, respectively. However, α-fibrinogenases in venoms hydrolyze the α subunit preferentially (sometimes exclusively), leading to

consumptive coagulopathies in envenomated patients (Mackessy 2010c). Both adult and juvenile *N. kaouthia* have venom α-fibrinogenases that hydrolyze the Aα subunit of fibrinogen, and both appear to catalyze hydrolysis at the same rate, with loss of the Aα-subunit within 30 minutes (Fig. 3). The β- and γ-subunits remained intact over the 60 minute time course. These results are consistent with effects of *N. n. karachiensis* venom on aPTT, PT and TT with platelet-poor plasma (Asad et al. 2012). Das et al. (2013) have also reported the degradation of Aα-chain of fibrinogen and anticoagulant activity of venom of *N. kaouthia* from northeast India. A very recent report indicates that a 66 kDa metalloproteinase purified from *N. kaouthia* venom specifically hydrolyzed the Aα subunit of fibrinogen and also hydrolyzed fibrin (Chanda et al. 2015). It is also possible that serine proteases could also be involved (Mackessy 1993a, b).

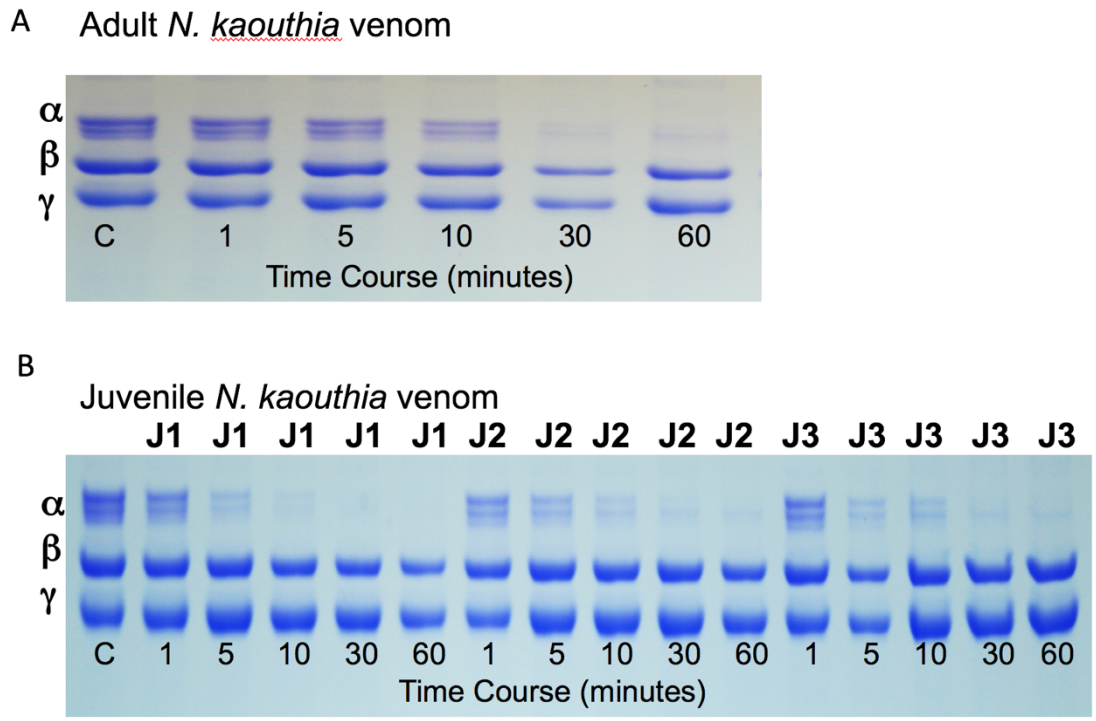


Figure 3. Adult and juvenile *Naja kaouthia* venom fibrinogen digest assay. Adult (A) and three different juvenile (B; J1-J3) venoms incubated with fibrinogen over a 60 minute time course. Fibrinogen subunits are labeled α, β, and γ. C, control: fibrinogen incubated in the absence of venom.

Chromatographic Venom Profiling, Toxin Purification, and Mass Spectrometry Identifications

Enzyme assays, 1D, and 2D gels all demonstrated a difference in abundance and activity of PLA_2 enzymes in juvenile and adult *N. kaouthia* venoms. Cation-exchange FPLC venom profiles for adult and juvenile cobras were generated to observe relative abundances of individual PLA_2 isoforms. FPLC was also used as the first step in the purification of the major proteins within adult *N. kaouthia* venom to determine if any abundant components in *N.*

kaouthia venom demonstrated significant toxicity towards a non-model organism (lizards).

FPLC chromatograms revealed several differences in protein abundance between adult and juvenile venoms. However, the number of peaks and elution time for each protein were essentially identical in both adult and juvenile *N. kaouthia* venoms (Fig. 4), indicating that the same proteins are present in both venoms but vary in abundance between adult and juvenile cobras; this type of intraspecific variation within cobras (variation in specific protein abundance between individuals) has been previously observed (Modahl et al. 2010).

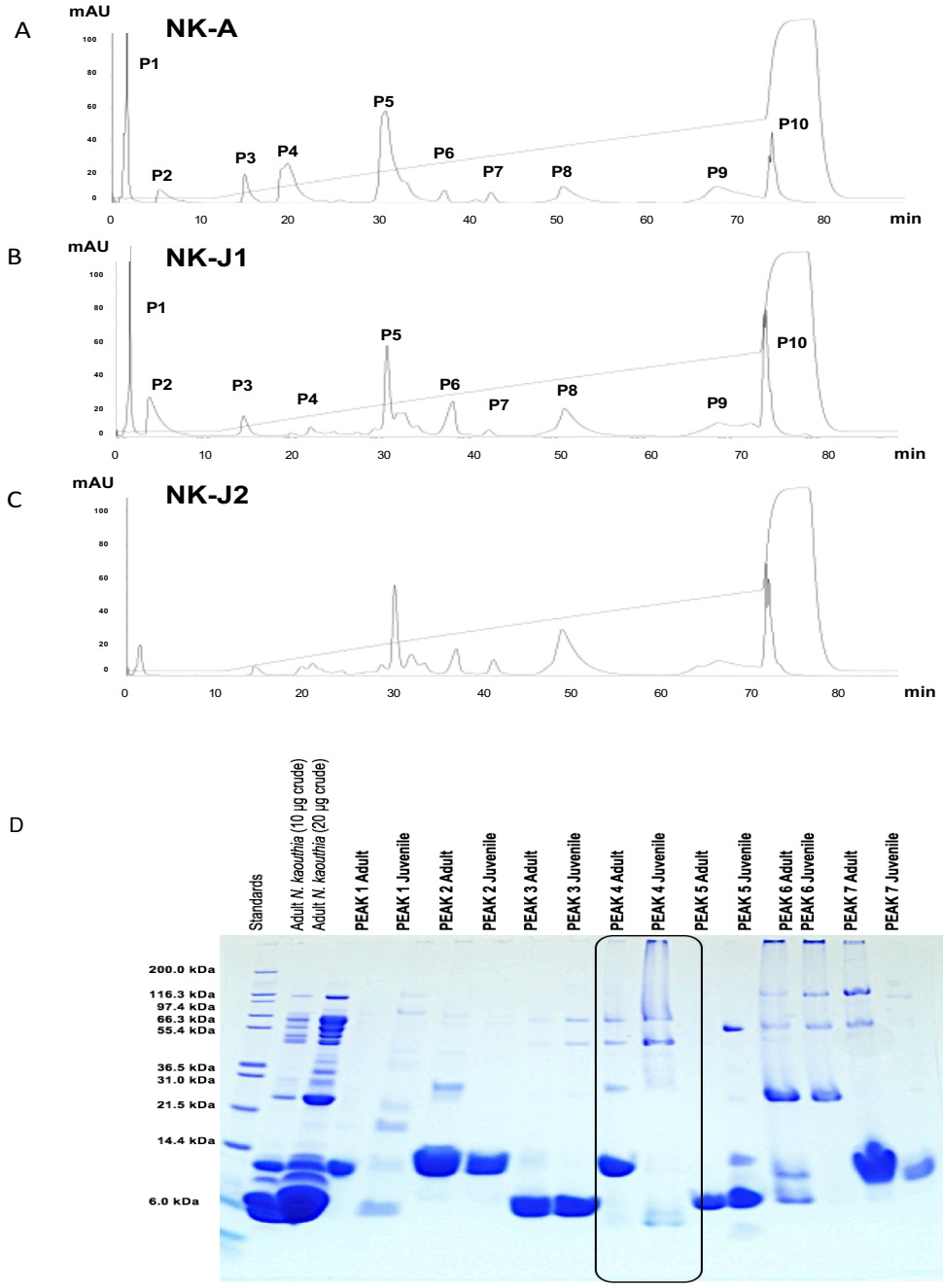


Figure 4. Adult and juvenile *Naja kaouthia* venom fractionation via cation-exchange FPLC. Adult *N. kaouthia* venom (3mg) was eluted with 20 mM MES and an NaCl gradient from 0 to 0.4 M over 70 minutes (A); the same methodology was used for several juvenile *N. kaouthia* venoms; two juvenile (NK-J1 and NK-J2) venom profiles shown (B and C, respectively). Peaks 1-7 from the adult and the NK-J1 juvenile venoms were subjected to reducing SDS-PAGE (Fig. 4D). Peak 4 (PLA_2) is highlighted within the chromatograms (adult fraction on left and juvenile fraction on right).

Protein components in each FPLC peak were first identified by their relative SDS-PAGE masses (Fig. 4D), and then identified from MALDI-TOF MS spectra. Most of the peaks contained proteins in the mass range of 6-8 kDa, which is consistent with SDS-PAGE (Fig. 1) and previously published *N. kaouthia* venom proteomes (Kulkeaw et al. 2007; Tan et al. 2015). This is the molecular mass range observed for 3FTxs, which make up a large majority of proteins in many elapid venoms, including *N. kaouthia* (Kulkeaw et al. 2007; Mackessy 2010b; Laustsen et al. 2015; Tan et al. 2015).

Of particular interest were proteins of 13-14 kDa, which is the mass range that has been reported for PLA_2 proteins in *N. kaouthia* venoms (Doley and Mukherjee 2003; Kulkeaw et al. 2007; Tan et al. 2015). It was observed that for peak four, there was an abundance of PLA_2s in the fractionated adult *N. kaouthia* venom, but the juvenile venom lacked this PLA_2 band (Fig. 4D). Peak four of the adult *N. kaouthia* FPLC chromatogram consisted of 11.4% of the total area of all peaks; for juvenile venoms, this peak consisted only of 1.4-4.9% of the total peak area. MALDI-TOF MS spectra revealed that the primary component of peak four for adult *N. kaouthia* venom is a 13,248 Da protein (PLA_2) and in the same peak in juvenile *N. kaouthia* venom there is instead a 6,841 Da protein (3FTx) as the primary component (Fig. 5). To confirm the identity of the 13,248 Da protein, PLA_2 assays showed activity of 0.0452 µmol/min/ml for fractions from peak four from adult *N. kaouthia* venom and 0.001 µmol/min/ml for peak four from juvenile venom.

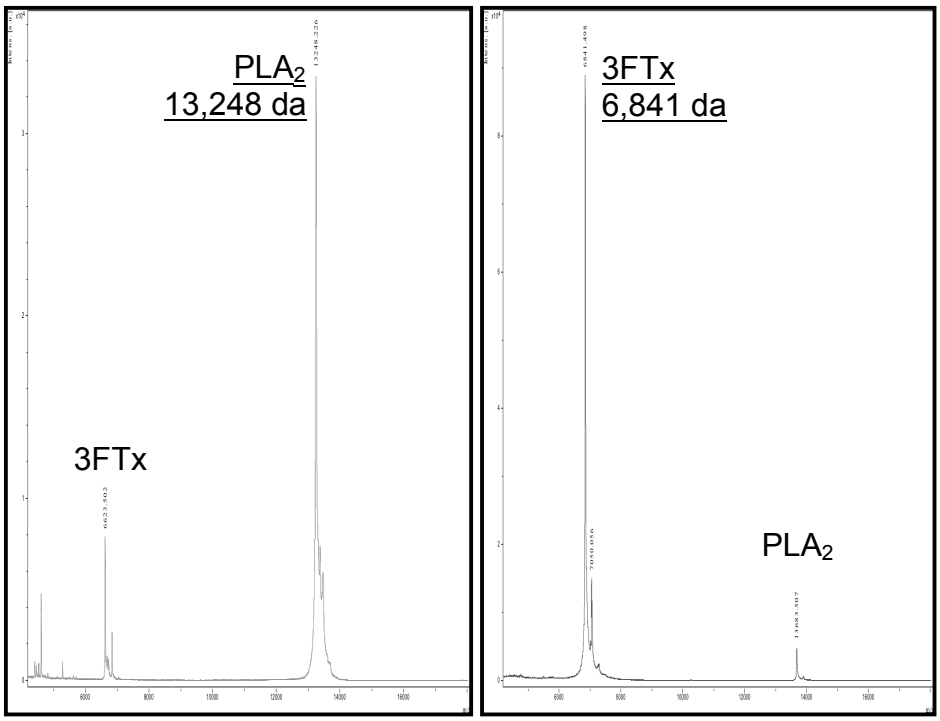


Figure 5. MALDI-TOF spectra for Peak 4 of *Naja kaouthia* venom cation-exchange FPLC fractionations. MALDI-TOF spectra of Mono S peak 4 of adult (A) and juvenile *N. kaouthia* venoms (B).

All major FPLC cation-exchange peaks from the adult *N. kaouthia* chromatogram above 20 mAU (except the first eluting peak of unbound proteins) were subjected to a second purification step using reversed-phase HPLC to purify the most abundant toxin within each major peak (Fig. 6). These FPLC peaks (peak four, five, eight, nine, and ten) comprised approx. 70% of the adult crude *N. kaouthia* venom composition (based on FPLC chromatogram calculated peak area). Of primary interest was the PLA_2 isoform(s) that was nearly absent from juvenile *N. kaouthia* venoms (peak four) and the most abundant weak non-conventional 3FTx from *N. kaouthia* venom.

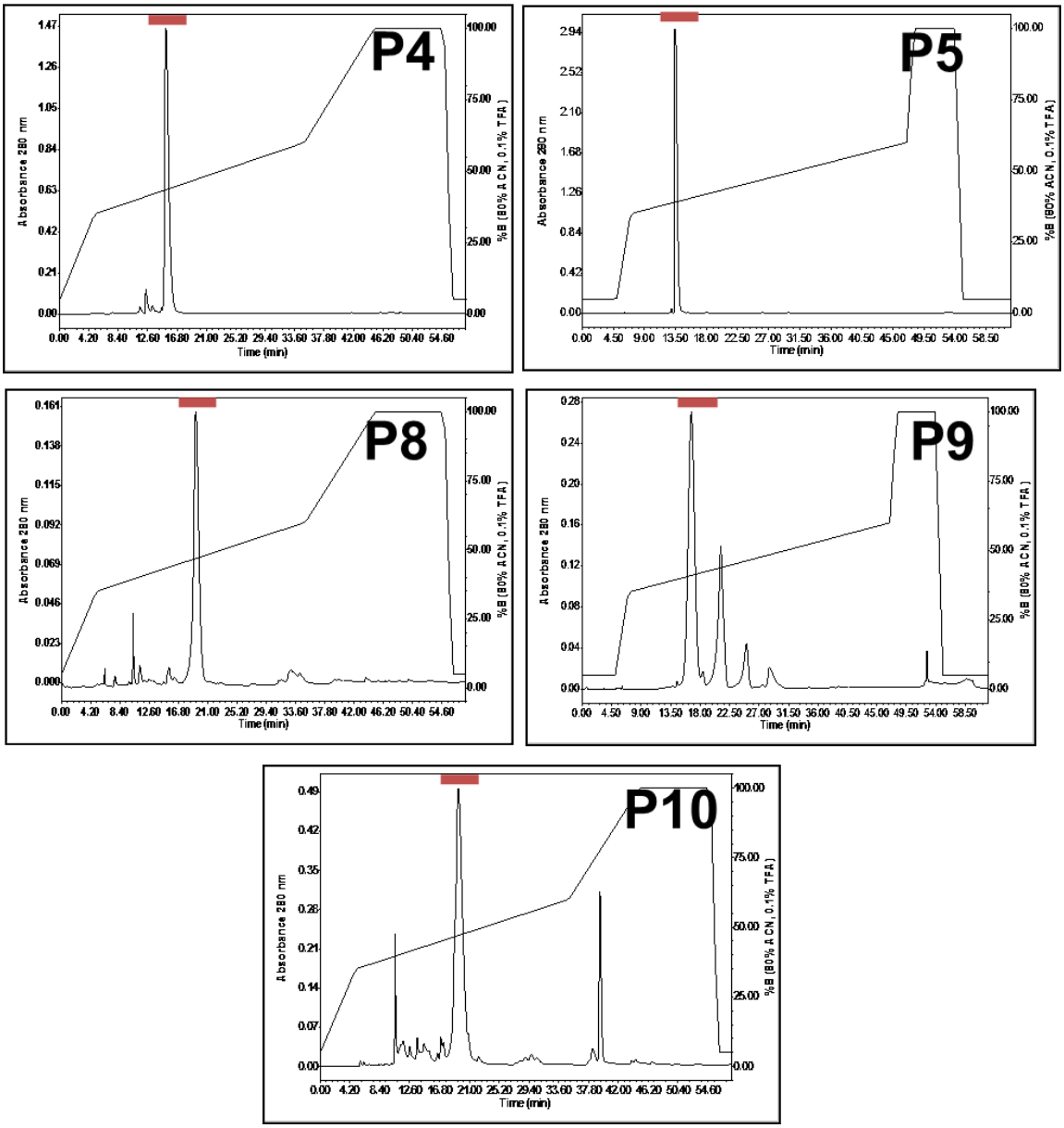


Figure 6. Major protein peaks (P4, P5, P8, P9, and P10) from FPLC Mono S cation-exchange fractionation of adult *N. kaouthia* venom were subjected to a reversed-phase HPLC final purification step. The most abundant purified toxin from each chromatogram (red bar) was identified via HPLC-nESI-LIT-Orbitrap MS/MS and used for lethal toxicity (LD_{50}) assays.

Purified toxin from the fourth peak (adult venom) was a PLA_2 isoform with a molecular mass of 13,248 Da (Figs. 5A, 6). Approximately 97% of the total MS/MS spectra (after normalization of spectra to protein molecular mass) were peptides from PLA_2 isoforms; a PLA_2 similar to natratoxin (PA2A_NAJAT) from *Naja atra* venom had the most abundant hits (Table 2). The fifth Mono S peak (Fig. 6) was a long-chain 3FTx, alpha-cobratoxin (alpha-elapitoxin; 3L21_NAJKA), with a molecular mass of 7,817 Da (94% of total spectra; Table 2). Cardiotoxic/cytotoxic 3FTxs (97%) were the most abundant toxins from the eighth peak, and cytotoxic 3FTxs (94%) were the most abundant toxins within the tenth peak (Fig. 6; Table 2). The ninth peak contained the most abundant weak non-conventional 3FTxs within *N. kaouthia* venom (58%). Peptide sequence coverage of each reported toxin ranged from 62-100%. The identities of the most abundant toxins within this study are also consistent with the previously published *N. kaouthia* (Thailand origin) venom proteome, in which the most abundant toxins were alpha-elapitoxin/long-chain neurotoxins (33.4%), cytotoxic/cardiotoxic 3FTxs (27.6%), PLA_2s (12.2%), weak neurotoxic 3FTxs (8.9%) and short-chain 3FTxs (7.7%); all other individual protein families accounted for less than 3% (Tan et al. 2015).

Table 2. Peptides and identities of purified toxins within major FPLC peaks.

Peak	%Total spectra	MS/MS-derived Peptide Sequences	NCBI/Uniprot Accessions	Venom protein
P4	97%	(K)NmIQcTVPSR(S) (K)GGSGTPVDDLDR(C) (R)ccQVHDNcYNEAEK(I) (R)ccQVHDNcYNEAEKISGcWPYFK(T) (K)ISGcWPYFK(T) (K)TYSYEcSQGTLTcK(G) (K)GGNNAcAAAVcDcDR(L) (K)TYSYEcSQGTLTcK(G) SWWDFADYGcYcGR LAAIcFAGAPYNNNNYNIDLK (K)GGSGTPVDDLDR(C) GDNDAcAAAVcDcDR ccQVHDNcYNEAEKISR SWWDFADYGcYcGR (K)GGSGTPVDDLDR(C) SWWDFADYGcYcGR LAAIcFAGAPYNDNNYNIDLK (R)ccQVHDNcYNEAEK(I) GGNDAcAAAVcDcDR LAAIcFAGAPYNDNNYNIDLKAR NMIQcTVPNRSWWDFADYGcYcGR GGNNAcAAAVcDcDRLAAIcFAGAPYND N-NYNIDLK GGSGTPVDDLDRccQIHDNcYNEAEK NMIQcTVPNR (K)GGNNAcAAAVcDcDR(L)	PA2A_NAJAT sp|P00596|PA2A1_ NAJKA Acidic phospholipase A2 sp|P15445|PA2A2_ NAJNA Acidic phospholipase A2 sp|Q9I900|PA2AD_ NAJSP Acidic phospholipase A2 sp|Q92084|PA2NA _NAJSP Neutral phospholipase A2	PLA_2
P5	94%	IRcFITPDITSKDcPNGHVcYTK(T) (K)TWcDAFcSIRGK(R) (R)GKRVDLGcAATcPTVK(T) (R)VDLGcAATcPTVK(T) (K)DcPNGHVcYTK(T) (R)GRRVDLGcAATcPTVK(T) (K)TGVDIQCCSTDNcNPFPTRP	3L21_NAJKA 3L23_NAJNA E2IU01_NAJAT 3L22_NAJNA	3FTx (long-chain)
P8	97%	LKcNKLIPIASK(T) (K)LIPIASKTcPAGK(N) (K)MFmMSDLTIPVK(R) (R)GcIDVcPKNSLLVK(Y) (K)NSLLVKYVccNTDRcN IRcFITPDITSKDcPNGHVcYTK(T) (K)DcPNGHVcYTK(T) (K)TWcDAFcSIR(G) (R)GKRVDLGcAATcPTVK(T) LKcNKLIPLAYK(T) (K)TcPAGKNLcYK(M) (K)RGcIDVcPK(N)	3SA8_NAJKA 3L23_NAJNA 3SA1_NAJNA	3FTx (cyto-toxin)

Table 2 continued.

P9	58%	EMIEccSTDKcNR LTcLNcPEMFcGK NGEKIcFKK (R)NGEKIcFK(K)	sp|P25679|3NO29_ NAJKA Weak toxin CM-9a 3NO2_NAJKA	3FTx (weak toxin)
P10	94%	(K)KFPLKIPIKR(G) (R)GcADNcPKNSALLK(Y) LKcHNTQLPFIYK(T) NSALLKYVccSTDKcN (K)YVccSTDKcN (T)LKcNKLVPLFYK(T) (K)LVPLFYKTcPAGK(N) (K)TcPAGKNLcYK(M) (K)MFMVATPKVPVKR(G) (K)RGcIDVcPK(S) (R)GcIDVcPKSSLLVK(Y) (K)SSLLVKYVccNTDRcN (K)YVccNTDRcN LKcNKLIPIASK(T) (K)LIPIASKTcPAGK(N) (K)MFmMSDLTIPVKR(G) (K)RGcIDVcPK(N) (K)NSLLVKYVccNTDRcN LKcNKLIPLAYK(T) (K)MFmVAAPK(V) (R)GcIDAcPK(N)	sp|P14541|3SOFH_ NAJKA Cytotoxin homolog 3SA0_NAJSP 3SA8_NAJKA 3SA5_NAJKA 3SA1_NAJNA	3FTx (cyto- toxin)

Lethal Toxicity (LD_{50})

Within cobra venoms, there are several different toxicity levels of 3FTxs: the “conventional” neurotoxins, with low murine LD_{50} values (~0.04-0.3 μg/g), and the “non-conventional” toxins, with higher murine LD_{50}s (~5-80 μg/g); (Nirthanan et al., 2003). However, the toxicity of these 3FTxs have only been determined using murine models, and many elapids exhibit non-mammalian prey preferences, often feeding on a diversity of prey, including lizards and other ectotherms (Luiselli et al., 1997, 2002). The major non-conventional weak 3FTx within *N. kaouthia* venom was of particular interest because other non-conventional 3FTxs are currently the only identified prey-specific toxins (Pawlak

et al., 2006; Pawlak et al., 2009; Heyborne and Mackessy, 2013). Non-conventional 3FTxs in rear-fanged snake venoms exhibit potent neurotoxicity towards non-murine species (lizards and birds; Pawlak et al., 2006; Pawlak et al., 2009; Heyborne and Mackessy, 2013). Because of these observations, all purified major non-conventional weak 3FTxs and other abundant toxins within *N. kaouthia* venom were tested for toxicity toward an ectothermic organism (*Hemidactylus frenatus*).

Adult and juvenile *N. kaouthia* crude venoms were evaluated using both lizards and mice to determine if a difference exists in toxicities towards non-model and model prey species. Differential venom toxicities towards different prey have been previously reported and linked to ontogenetic dietary shifts (Mackessy 1988; Mackessy et al. 2006). The mouse i.p. LD_{50} value for adult *N. kaouthia* venom was 0.61 (95% confidence interval - 0.43-0.80) µg/g (Table 3), which is within the range of what has been previously reported (0.70 ± 0.09 µg/g) (Mukherjee and Maity 2002); for juvenile *N. kaouthia* venom, the LD_{50} was 0.75 (95% CI - 0.60 - 0.91) µg/g. Because the lower and upper 95% confidence intervals for mouse LD_{50} values show extensive overlap, it was concluded that no difference in mammalian toxicity exists between adult and juvenile cobra venoms. However, the House Gecko i.p. LD_{50} value for adult *N. kaouthia* venom was 1.02 (0.88 - 1.16) µg/g and that for juvenile *N. kaouthia* venom was found to be 1.29 (1.26 – 1.32) µg/g (Table 3). In this case, the lower and upper 95% confidence intervals for House Gecko LD_{50} values do not overlap, indicating that adult *N. kaouthia* venoms are slightly more toxic toward lizards. Phospholipase

enzymes, which are in greater abundance in adult *N. kaouthia* venoms, are among the more toxic and pharmacologically active components of snake venoms (Mackessy 2010b). Higher PLA_2 enzyme concentrations, especially PLA_2s that covalently interact with low molecular mass weak-neurotoxin like peptides (kaouthiotoxins), are often related to increased toxicity (Mukherjee and Maity 2002; Mukherjee 2010). It is therefore possible that the greater abundance of PLA_2 enzymes within adult *N. kaouthia* venoms results in somewhat more toxic adult cobra venoms towards House Geckos.

Table 3. Lethal toxicity (LD_{50}) values for *Naja kaouthia* crude venom (adult and juvenile) using model and non-model organisms for toxicity determination.

Species	Adult *N. kaouthia* venom	Juvenile *N. kaouthia* venom
Mice (*Mus musculus*)	0.614 ± 0.182 µg/g i.p	0.754 ± 0.157 µg/g i.p
House Geckos (*Hemidactylus frenatus*)	1.023 ± 0.139 µg/g i.p	1.290 ± 0.032 µg/g i.p

i.p. = intraperitoneal

Toxicity assays were also performed using HPLC-purified abundant venom proteins from adult *N. kaouthia* venom (Table 4) to determine if any toxins exhibited prey-specific toxicity. These toxins included PLA_2s, long-chain 3FTxs (α-cobratoxin), weak non-conventional 3FTxs, and the two primary cardiotoxic/cytotoxic 3FTxs. With the exception of the long-chain α-cobratoxin (alpha-elapitoxin; 3L21_NAJKA), intraperitoneal (i.p.) LD_{50} values were all above 5 µg/g (the highest dose tested) for lizards (*Hemidactylus frenatus*) (Table 4).

Published toxicity values (LD_{50}s) observed in mice for the major cytotoxins within *N. kaouthia* venom range from 1.2 -1.48 µg/g intravenous (i.v.) and 2.25 µg/g intraperitoneal (i.p.) (Joubert and Taljaard 1980a; Ohkura et al. 1988). Therefore, *N. kaouthia* cytotoxins were less toxic towards lizards than toward mice (Table 4). The LD_{50} value of α-cobratoxin fell below 0.1 µg/g (the lowest dose tested), which is the same as previous reported values for this toxin in mice (Karlsson and Eaker 1972). These results demonstrate that unlike venom from the rear-fanged *B. irregularis*, which contains an abundant, prey-specific (lizards and birds), dimeric neurotoxic 3FTx (Pawlak et al., 2009), *N. kaouthia* venom has an abundant neurotoxic 3FTx that can bind effectively (and lethally) to the nicotinic acetylcholine receptors of a variety of prey types (both endothermic and ectothermic). As a generalist predator, *N. kaouthia* consume a variety of prey, and a “generalist” 3FTx that can incapacitate a large diversity of prey types is the main lethal toxin that has evolved in this venom. However, it is important to note that these results are only for one lizard species (*Hemidactylus frenatus*); other lizard species or other ectothermic species might respond differently to *N. kaouthia* venom toxins. The *H. frenatus* used for this study were chosen because they are a species that is sympatric with *N. kaouthia* and therefore a potential prey item; they were also commercially available in the numbers needed for toxicity determinations.

Table 4. Lethal toxicity (LD_{50}) of purified *N. kaouthia* major venom toxins.

Peak / Toxin	Accession Number	LD_{50} Values for Lizards (*Hemidactylus frenatus*)	LD_{50} Values for Mice (*Mus musculus*)
P4: Acidic PLA_2, similar to natratoxin	PA2A_NAJAT	LD_{50} > 5 µg/g i.p.	LD_{50} > 5 µg/g i.v. (Joubert and Taljaard 1980b)
P5: α-elapitoxin	3L21_NAJKA	LD_{50} < 0.1 µg/g i.p.	LD_{50} < 0.1 µg/g i.v. (Karlsson 1973)
P8: Cardio-cytotoxin	3SA8_NAJKA	LD_{50} > 5 µg/g i.p.	LD_{50} = 2.25 µg/g i.p. (Joubert and Taljaard 1980a)
P9: Non-conventional weak 3FTx	3NO29_NAJKA	LD_{50} > 5 µg/g i.p.	LD_{50} > 5 µg/g i.v. (Joubert and Taljaard 1980c)
P10: Cytotoxin	3SOFH_NAJKA	LD_{50} > 5 µg/g i.p.	LD_{50} = 2.25 µg/g i.p. (Joubert and Taljaard 1980a)

i.p. = intraperitoneal; i.v. = intravenous

Because overall venom toxicity towards mice and lizards appears to result from α-cobratoxin, it is possible that the significant difference in PLA_2 activity between adult and juvenile *N. kaouthia* venoms primarily serves a pre-digestive role, rather than a major role in lethal toxicity. Snakes are gape-limited feeders, and juveniles are restricted to smaller prey. Larger prey are consumed by adult snakes, and the increase in PLA_2 enzymes in adult *N. kaouthia* venom may facilitate digestion of these larger prey by promoting myonecrosis. The biological roles of enzymatic venom proteins within viperids (particularly snake venom metalloproteinases) have been suggested to include aiding digestion (Lomonte et

al., 2009; Mackessy, 2010a), and there may be functional convergence toward a pre-digestive role for some PLA_2s within elapid venoms.

Conclusions

Ontogenetic differences in composition of adult and juvenile *N. kaouthia* venoms were shown to exist, but these differences are not as pronounced as is typical of many viperid venoms. One- and two-dimensional SDS-PAGE results indicate a greater abundance of cobra venom factor and CRiSP isoforms in juvenile cobra venoms, and adult cobra venoms show a higher abundance of PLA_2 enzymes, including the least acidic PLA_2 isoforms that are completely absent from some juvenile venoms. Enzyme activities matched these results, with significantly lower PLA_2 activities observed in juvenile venoms. There was also a significant difference in phosphodiesterase and acetylcholinesterase activities between adult and juvenile venoms. Metalloproteinase, L-amino acid oxidase, and fibrinogen digest assays revealed no significant differences in activities between adult and juvenile venoms. Adult *N. kaouthia* crude venom does appear to be slightly more toxic towards House Geckos, but adult and juvenile crude *N. kaouthia* venoms were equally toxic towards mice.

Juvenile and adult cobras both consume endothermic and ectothermic prey and do not appear to shift in prey preferences (other than size) as they age (Chaitae 2000). Although *N. kaouthia* also prey upon ectothermic prey (lizards and snakes), none of the non-conventional 3FTxs showed prey-specific effects, so our initial hypothesis was not supported. The main neurotoxic 3FTx within *N. kaouthia* venom, α-cobratoxin (α-elapitoxin), exhibits the same potent toxicity

towards both mammalian and lizard prey. The presence of α-cobratoxin in both adult and juvenile cobra venoms represents a venom compositional strategy wherein a single potent toxin effectively immobilizes a variety of prey types encountered across snake life history stages. Future studies could evaluate possible ontogenetic venom variation in other cobra species that do have a significant age-related dietary shift. Similarly, investigation of non-conventional 3FTxs from cobra species which prey more extensively on non-mammalian species, such as *Ophiophagus hannah*, which is largely a snake specialist, could reveal novel prey-specific toxins. Because venoms evolved as trophic adaptations that facilitate prey handling, it is important to explore intraspecific variation at several different levels, including overall venom protein composition, enzymatic activities, and toxicity to fundamentally different prey. Combined with proteomic and transcriptomic data, this functional data can provide deeper insight into the evolution of venom diversification and its relation to trophic habits of snakes.

CHAPTER III

TRANSCRIPTOME-FACILITATED VENOMICS OF THE AMAZON PUFFING SNAKE (SPILOTES {PSEUSTES] SULPHUREUS; COLUBRIDAE): EXPLORING THE EVOLUTION AND PREY-SPECIFIC TOXICITY OF THREE-FINGER TOXINS

Cassandra M. Modahl[a], Seth E. Frietze[a,b], and Stephen P. Mackessy[a]
[a]School of Biological Sciences, University of Northern Colorado, 501 20th St., Greeley, CO 80639-0017, USA
[b]Department of Medical Laboratory and Radiation Sciences, University of Vermont, 302 Rowell, Burlington VT 05405, USA

Abstract

Snakebite from rear-fanged venomous snakes rarely result in systemic envenomation for humans; however, several species of rear-fanged snakes do produce toxic venom, including *Dispholidus typus*, *Thelotornis* sp., yet venoms from most rear-fanged snake species remain unstudied. To characterize the venom composition of the rear-fanged *Pseustes sulphureus* (Amazon Puffing Snake), we applied a transcriptomic and venomic proteomic profiling approach, combined with enzymatic assays. The venom gland transcriptome contains full-length transcripts for 13 venom protein superfamilies, with the most abundant superfamilies consisting of three-finger toxins (60% of toxin reads), C-type lectins (11%) and ficolins (8.4%). However, ficolins were not detected in the venom proteome. The venom proteome consisted primarily of three-finger toxins, cysteine-rich secretory proteins, PIII metalloproteinases, L-amino acid oxidases,

and PLA_2 inhibitors. Three-finger toxins were the most abundant proteins within the venom, making up approximately 91-92%. The murine (*Mus musculus*) lethal dose (LD_{50}) for crude *P. sulphureus* venom was 2.56 µg/g, and for House Geckos (*Hemidactylus frenatus*) it was even lower (1.01 µg/g). The three most abundant 3FTxs (sulmotoxin A, B, and C) were purified and exhibited differences in prey toxicity. Sulmotoxin A, a heterodimeric 3FTx complex, was highly toxic (LD_{50} = 0.22 µg/g) toward House Geckos and non-toxic (LD_{50} > 5 µg/g) toward NSA mice. In contrast, sulmotoxin B, a monomeric 3FTx, was toxic toward NSA mice and non-toxic towards House Geckos. The amino acid sequences of sulmotoxin A and sulmotoxin B varied in the second structural loop, which is responsible for receptor binding, and differences observed may be responsible for the differences in taxon specificity.

Biological significance

The newly described venom gland transcriptome and proteome of *P. sulphureus* reveals a rear-fanged snake venom composed largely of 3FTxs and that quite nearly lacks any detectable enzymatic activity. This high abundance of 3FTxs and absence of many enzymes is a venom phenotype that resembles many front-fanged elapid snakes (cobras, kraits, mambas, etc.). The two most abundant 3FTxs within this venom display differential and prey-specific toxicity, with sulmotoxin A exhibiting reptile-specific toxicity and the sulmotoxin B mammal-specific toxicity. The sulmotoxins exhibited such low similarity to other 3FTxs within public databases utilized for MS/MS spectral matching that a reanalysis of the proteomic data by interrogating the *P. sulphureus* venom gland

transcriptome was necessary to identify with confidence that the sulmotoxins belong to the 3FTx venom superfamily. This represents an important limitation for characterizing rear-fanged snake venoms proteomically, because there is a lack of representation of these sequences in current databases and therefore tryptic peptide matching approaches perform poorly. By obtaining the venom gland transcript sequences, including those for sulmotoxins, the corresponding purified proteins found in the gland could be matched to provide the amino acid sequences responsible for the specific binding. Results provided insight into potential amino acids responsible for the specific binding and prey-specific toxicity observed. This is the first time 3FTxs that target different prey have been characterized in a single venom, emphasizing how diversification of 3FTxs has resulted in a simple venom system with maximal flexibility for capture and incapacitation of diverse prey. This whole systems approach toward characterizing components of a snake venom (venom gland transcriptome, venom proteome, biochemical and biological assays) illustrates that an integrated approach toward understanding snake venom complexity results in a deeper understanding of the evolution of venoms. Further, the relative minimalist arsenal of some rear-fanged snakes is more tractable to a combined high- and low-throughput series of analyses, demonstrating that understanding rear-fanged snake venom composition is central toward understanding venom evolution in the advanced snakes.

Introduction

Venomous snakes dangerous to humans belong almost exclusively to the front-fanged Viperidae (vipers and pit vipers) and Elapidae (cobras, mambas, kraits, and relatives). Their venoms have been the primary focus of most venom research because of the medical relevance of bites from these snakes, and human morbidity and mortality resulting from these bites is currently recognized as a major neglected tropical disease (Gutiérrez et al. 2010; Harrison et al. 2011). However, several rear-fanged venomous snake species have also been responsible for human mortalities, including fatal envenomations of noted herpetologists Karl Schmidt (by *Dispholidus typus*) and Robert Mertens (by *Thelotornis capensis*), and many rear-fanged venomous snakes possess venoms with currently unstudied toxicities (Mackessy 2002; Zelanis et al. 2010; Weinstein et al. 2011; Modahl et al. 2015). Characterization of rear-fanged snake venoms is therefore important to identify snakes that could cause clinically-significant human envenomations. Of greater biological interest is the potential for these species to harbor novel toxins, and because rear-fanged venomous snakes are distantly related to front-fanged snakes and display unique evolutionary trajectories, their venoms are of critical importance for understanding the evolution of venoms among the advanced snakes (superfamily Colubroidea). For example, at present, only rear-fanged species are known to produce three-finger toxins (3FTxs) having prey-specific effects (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013), but it is unknown how widely occurring this phenomenon is among snakes.

Prey-specific toxins have been isolated from two species of rear-fanged *Boiga sp.* (Cat snakes) and *Oxybelis fulgidus* (Green Vinesnake). All three identified prey-specific toxins were found to be lethal towards lizards and/or birds, with lethal toxicities (LD_{50}s) in lizards/birds ranging from 0.1-0.3 μg/g; all were non-lethal towards mammals (LD_{50} >5 μg/g) (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013). In addition, all identified prey-specific toxins have belonged to the three-finger toxin superfamily, and one forms a novel heterodimeric complex (Pawlak et al. 2009).

Three-finger toxins (3FTx) comprise a non-enzymatic superfamily of venom proteins that are common components of many snake venoms, and they share an identical structural scaffold consisting of three β-stranded loops ("fingers") stabilized by four conserved disulfide bridges (Kini and Doley 2010; Mackessy 2010b). Three-finger toxins are an example of a venom protein superfamily in which all members possess a nearly identical three dimensional structure but exhibit a large diversity of biological activities, including neurotoxic, cardiotoxic, cytotoxic, hemorrhagic, or tissue necrotizing effects (Kini and Doley 2010). These toxins are most abundant in elapid venoms where they are responsible for the neurotoxicity and respiratory paralysis that can result in death from envenomations by these snakes; these toxins have also been identified in several rear-fanged venomous snakes, but they do not appear to impart lethal (mammalian) toxicities to these venoms (Fry et al. 2003a; Fry et al. 2003c; Pawlak et al. 2006; Pawlak et al. 2009; Mackessy 2010b; Weinstein et al. 2011; Fry et al. 2012; Heyborne and Mackessy 2013). Several members of the genus

Boiga (Asian Catsnakes) produce venoms with some of the highest abundances of three-finger toxins currently known in rear-fanged snake venoms (Fry et al. 2003c; Mackessy et al. 2006; McGivern et al. 2014), and one species (*Boiga irregularis*) is capable of causing significant human morbidity following extended contact bites (Weinstein et al. 2011). These known examples, coupled with the fact that most species' venoms are wholly unknown, prompted us to investigate the venom produced by other large species.

The rear-fanged *Pseustes sulphureus* (Amazon Puffing Snake) is native to South America. These snakes are semi-arboreal and are among the largest of the advanced snakes, reaching up to 2.7m (Lopez and Maxson 1995). At present, nothing is known about their venom production ability or venom composition, but being a larger snake and becoming common in the herpetocultural trade, they have the potential for producing high venom yields which could result in serious envenomations. We present here the first complete compositional study of *P. sulphureus* venom and suggest that its "harmless" reputation is likely undeserved.

A major hurdle to characterizing the venoms of rear-fanged snakes is the lack of rear-fanged snake venom protein sequences within public databases and the presence of novel protein families within their venoms (Lumsden et al. 2007; OmPraba et al. 2010; Ching et al. 2012). The emergence of "-omic" technologies in recent years has revolutionized venom research by integrating high-throughput genomic, transcriptomic, and proteomic data to generate comprehensive molecular venom profiles (Calvete 2013; Calvete 2014). A comprehensive proteomics

approach, with tandem mass spectrometry (MS/MS) peptide sequencing of separated venom components (usually by high performance liquid chromatography and/or two dimensional gel electrophoresis) combined with a species-specific venom gland transcriptome, has provided the most complete venom compositional coverage (Wagstaff et al. 2009; Calvete 2014; McGivern et al. 2014; Sunagar et al. 2014). However, MS/MS identification of proteins relying on online sequence databases can overlook unique isoform variations and therefore fail to recognize novel venom proteins because only small tryptic peptide fragments, which may contain unique sequences, are used for protein identification. This is a particularly acute problem when analyzing snake venoms, which characteristically have limited protein families with multiple structural isoforms with diverse activities. By generating a complementary transcriptome for a given species, MS/MS peptide sequences can be precisely matched to the corresponding transcript, and the translated transcripts provide full protein sequences, avoiding labor-intensive and expensive methodologies such as N-terminal sequencing and MS/MS-based *de novo* sequencing of many peptide fragments.

This combined transcriptomic and proteomic approach was found to be especially useful to characterize the currently unstudied venom from *P. sulphureus,* providing a comprehensive view of the venom gland transcriptome and venom proteome. Structural analyses were complemented with biochemical and biological assays of venom and venom components, providing a detailed evaluation of the evolution and biological roles of venom components, particularly the three-finger toxins within *P. sulphureus* venom. As a trophic adaptation, venom has the potential

to provide insights into unique predator and prey evolutionary dynamics. Further, venom protein specificity for receptors and ion channels can provide insight for molecular structure and function studies, as well as for development of potential therapeutics (Nirthanan and Gwee 2004; Yee et al. 2004; Kini and Doley 2010).

Materials and Materials

Materials

TRIzol reagent was purchased from Life Technologies (San Diego, CA, U.S.A.). cDNA library sequencing preparation kits, including the KAPA Stranded mRNA-Seq kit and KAPA Library Quantification Kit (Illumina® platforms) were purchased from KAPA Biosystems (Boston, MA, U.S.A.). Agencourt AMPure XP reagent from Beckman Coulter, Inc. (Brea, CA, U.S.A) was also purchased and used during cDNA library preparation procedures. Novex Mark 12 unstained molecular mass standards, MES running buffer, LDS sample buffer, and precast 12% Bis-Tris NuPAGE electrophoretic gels were obtained from Life Technologies (San Diego, CA, U.S.A.). DeStreak rehydration solution, IPG pH 3-11 buffer, and the IPGphor electrofocusing system were purchased from GE Healthcare (Pittsburgh, PA, U.S.A.). Pierce BCA protein assay kit was obtained from Thermo Fisher Scientific (Rockford, IL, U.S.A). Phospholipase A_2 assay kit was purchased from Cayman Chemical Co (Ann Arbor, MI, U.S.A.). Azocasein, dithiothreitol, iodoacetamide, acetylthiocholine iodide, bis-p-nitrophenylphosphate, L-kynurenine, benzoyl-Phe-Val-Arg-paranitroaniline substrate, benzoyl-Pro-Phe-Arg-paranitroaniline substrate, human fibrinogen, and all other reagents (analytical grade or better), were obtained from Sigma-

Aldrich (St. Louis, MO, U.S.A.). All reagents used for molecular work were certified nuclease-free.

Snake Venom and Venom Gland Tissue Collection

Venom was manually extracted from all rear-fanged snakes used in this study (*Boiga cynodon, B. irregularis, B. dendrophila, B. nigriceps, Pseustes sulphureus, Trimorphodon biscutatus lambda, Oxybelis fulgidus, Ahaetulla prasina, Lieoheterodon madagascariensis*, *Thrasops jacksonii, Rhamphiophis oxyrhynchus, Tantilla cucullata, Alsophis portoricensis, Psammophis schokari* and *Thelotornis kirtlandii*). Rear-fanged snakes were maintained in the University of Northern Colorado Animal Resource Facility in accordance with UNC-IACUC protocol #9204. Three Amazon Puffing snakes (*P. sulphureus*) originating from Suriname, South America were recent imports (Bushmaster Reptiles). All *P. sulphureus* were adults, weighing over 500 grams and measuring over 1,500 mm snout-to-vent lengths. Rear-fanged snakes were extracted using the method of Hill and Mackessy (1997) with subcutaneous injections of ketamine-HCl (20-30 mg/kg) followed by pilocarpine-HCl (6 mg/kg) (Hill and Mackessy 1997). Collected venom from each individual (venom was not pooled) was centrifuged at 9,500 x g for 5 minutes, frozen at -80°C, lyophilized, and stored at -20°C until use. Venom gland tissue (Fig. 7) was dissected from one *P. sulphureus* within 12 hours of death and was added to TRIzol reagent for immediate RNA isolation.

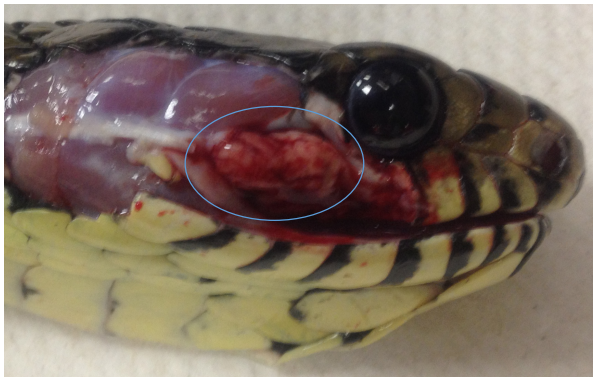


Figure 7. *Pseustes sulphureus* venom gland structure. Like most rear-fanged colubrids, it is located posterior and ventral to the eye.

Enzyme Assays

Pierce BCA protein assay kit (Thermo Fisher Scientific, Rockford, IL, U.S.A) using BSA as standard was used to determine concentrations for crude venom and individual purified proteins for all assays described. Metalloproteinase activity was determined using azocasein as a substrate and 20 µg of crude *P. sulphureus* venom (Aird and da Silva 1991). This activity was expressed as $\Delta A_{342\ nm}$/min/ per mg venom protein. L-amino acid oxidase activity was assayed according to Weisbach (1961) with 20 µg of crude *P. sulphureus* venom, and the activity was expressed as nmol of product formed/min/mg protein (Weisbach et al. 1961). Acetylcholinesterase activity was determined using 15 µg crude *P. sulphureus* venom, incubated with the acetylthiocholine iodide substrate in a cuvette at 37°C (Ellman et al. 1961). Absorbance at 412 nm was taken every 10 seconds for ten minutes and the linear portion of the graph produced was used to calculate activity as µmole product formed/minute/mg venom protein. Phosphodiesterase activity was assayed with 20 µg crude *P. sulphureus* venom using 1 mM bis-p-nitrophenylphosphate as a substrate, following the protocol

developed by Laskowski (1980); activity reported as $\Delta A_{400\ nm}$/min/mg protein. Phospholipase A_2 activity was determined using a commercially available kit (Cayman Chemical Co.) as described by the manufacturer, using 2 µg crude *P. sulphureus* venom in 200 µL total volume. Absorbance was measured at 414 nm every minute for five minutes and activity was reported as µmol product formed/minute/mg protein. Thrombin-like and kallikrein-like serine proteinase activity was assayed according to Mackessy (1993) with benzoyl-Phe-Val-Arg-paranitroaniline substrate (thrombin-like activity) and benzoyl-Pro-Phe-Arg-paranitroaniline substrate (kallikrein-like activity) (Mackessy 1993b). The substrates were incubated with 20 µg crude *P. sulphureus* venom and activity recorded as nmol product formed/minute/mg protein. All enzyme assays described above were performed in triplicate. Fibrinogenase activity was determined using 20 µg of crude *P. sulphureus* venom incubated with human fibrinogen (final concentration 0.5 mg/mL) at 37 °C in a total volume of 200µL for periods of 0, 1, 5, 10, 30, and 60 minutes (Ouyang and Huang 1979). Twenty µL of this reaction mixture was removed at each time point and mixed with an equal volume of 4% SDS and 5% 2-mercaptoethanol, then heated in boiling water for 10 minutes. Five µL aliquots were combined with 2x LDS buffer, electrophoresed on a 12% NuPAGE Bis-Tris gel, stained with Coomassie Brilliant Blue overnight, destained for two hours, and imaged.

One-Dimensional Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed using NuPage 12% Bis-Tris mini gels (Invitrogen, Inc., U.S.A) under reduced (50 mM dithiothreitol) and non-reduced conditions. Samples and buffers were prepared according to the manufacturer. Crude venom (20 µg) from each rear-fanged venomous snake and 5 µL of Novex Mark 12 unstained mass standard (Invitrogen, Inc., U.S.A) were added to lanes. The gel was run at 160 V, stained with Coomassie Brilliant Blue overnight, destained (50/40/10, v/v, ddH_2O, methanol and glacial acetic acid) for two hours, and scanned.

Two-Dimensional Polyacrylamide Gel Electrophoresis (2D SDS-PAGE)

Crude venom from *P. sulphureus* (200 µg) was added to 135 µL DeStreak rehydration solution containing 18 mM DTT and 0.5% IPG pH 3-11 buffer (GE Healthcare, Pittsburgh, PA, U.S.A). This solution was added to the ceramic holder, overlain with a 7 cm IPG pH 3-11 strip and then mineral oil, and placed on an IPGphor Electrofocusing System (GE Healthcare, Pittsburgh, PA, U.S.A). The following isoelectric focusing steps were performed: rehydration for 12 hours, 500 V for 30 minutes, 1000 V for 30 minutes, 5000 V for 3 hours, and a 40 V hold. For the second dimension electrophoresis, the focused IPG strip was equilibrated in a reduction buffer (I mL 4X LDS buffer, 400 µL 150 mM DTT [final concentration 15 mM], and 2.6 mL ddH_2o) at ambient temperature for 10 minutes and then equilibrated in an alkylation buffer (1 mL 4X LDS, 3 mL ddH_2O, and a

final concentration of 60 mM iodoacetamide) at ambient temperature for 10 minutes. SDS-PAGE was carried out using NuPage 12% Bis-Tris 2D well gels following the same procedure as described above for 1D SDS-PAGE.

High Performance Liquid Chromatography (HPLC) Purification of *Pseustes sulphureus* Venom Proteins

Size exclusion HPLC fractionation of crude *P. sulphureus* venom (3-8 mg; filtered with a 0.45 µm syringe tip filter) was performed on a Waters HPLC system operating under Empower software with a Yarra 3µm SEC-2000 300 x 7.80 mm column (Phenomenex, Torrance, CA, U.S.A.) and a flow rate of 150 µL per minute of 25 mM HEPES buffer, pH 6.8, containing 100 mM NaCl and 5 mM $CaCl_2$, for 120 minutes. Peaks 2, 3, 4, and 6 of the size exclusion chromatogram were then subjected to reversed-phase HPLC using the same Waters HPLC system with a Jupiter 5 µm C18 250 x 4.60 mm column (Phenomenex, Torrance, CA, U.S.A.). The RP-HPLC column was equilibrated with 0.1% trifluoroacetic acid (TFA) and proteins were eluted using a 30-45% linear gradient of acetonitrile (80% in 0.1% TFA) over 25 minutes at a flow rate of 1 mL/min. Elution of proteins was monitored at 280 and 220 nm. Purified proteins were then lyophilized and stored at -20 °C until use.

Separation of Three-Finger Toxin (Sulmotoxin A) Subunits

Purified protein (approximately 100 µg) was treated with 250 mM dithiothreitol (DTT) dissolved in an ammonium bicarbonate solution (50 mM

NH_4HCO_3) and incubated at 50°C for 30 minutes. Iodoacetamide (IAA; 275 mM) was added, and the sample then incubated an additional 30 minutes at ambient temperature in the dark. The reduced and alkylated protein was subjected to RP-HPLC using a Jupiter 5 µm C18 250 x 4.60 mm column (Phenomenex, Torrance, CA, U.S.A.) attached to a Waters HPLC system. The column was equilibrated with 0.1% TFA and proteins were eluted using a 30-60% linear gradient of acetonitrile (80% in 0.1% TFA) over 50 minutes at a flow rate of 1 mL/min. Elution of proteins was monitored at 280 nm and 220 nm.

Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight Mass Spectrometry (MALDI-TOF)

Approximately 1 µg of reversed-phased HPLC purified 3FTxs or separated subunits was dissolved in 50% acetonitrile, 0.1% TFA in ddH_2O and mixed with an equal volume of sinapinic acid matrix (10 mg/mL in 50% ACN, 0.1% TFA); 4 µg of crude venom was also mixed with matrix and spotted onto a MALDI target. The molecular mass of each 3FTx and subunit was determined from spectra produced by a Bruker Microflex mass spectrometer (Metabolomics and Proteomics Facility, Colorado State University, Fort Collins, CO, U.S.A) operating in positive linear mode.

Lethal Dose (LD_{50}) Toxicity Assays

Crude *P. sulphureus* venom toxicity and the toxicity of the three most abundant three-finger toxins observed within *P. sulphureus* venom were evaluated with LD_{50} assays using a non-model species, House Gecko

(*Hemidactylus frenatus*), and a standard model species, NSA mouse (*Mus musculus*). House Geckos were obtained from Bushmaster Reptiles (Longmont, CO, U.S.A.) and NSA mice were bred in the UNC Animal Resource Facility (UNC IACUC protocol #9401). Doses used for crude venom were 0.5 µg/g, 1 µg/g, 3 µg/g, and 5 µg/g for geckos, and 0.5 µg/g, 1 µg/g, 2 µg/g, 2.5 µg/g, 3 µg/g, and 5 µg/g for mice. Doses for individual purified three-finger toxins were 0.1 µg/g, 0.5 µg/g, 0.8 µg/g, 1 µg/g, 3 µg/g, and 5 µg/g for geckos, and 0.3 µg/g, 1 µg/g, 3 µg/g, and 5 µg/g for mice. All doses were injected intraperitoneally in sterile 0.9% saline and three animals per dose were used. Three mice and three lizards were injected with only 0.9% saline as controls. Doses were adjusted to individual animal body masses, with lethality expressed as µg venom per gram body mass producing 50% mortality after 24 hours (Reed and Muench 1938). The nonparametric Spearman-Karber method was used for LD_{50} value estimations and determination of 95% confidence intervals (Finney 1978). All procedures were reviewed and approved by the UNC IACUC (protocol 1504D-SM-SMLBirds-18).

RNA Isolation, cDNA Library Preparation, and Next-Generation Sequencing

RNA isolation was completed following the manufacture's protocol for the TRIzol reagent with an overnight -20 °C incubation in 300 µL 100% ethanol, with 40 µL 3 M sodium acetate to increased RNA yields; total RNA obtained was resuspended in 16 µL nuclease-free H_2O. Poly-A+ RNA was selected from 4 µg of total RNA with oligo-dT beads using the KAPA Stranded mRNA-Seq kit (KAPA

Biosystems, Boston, MA, U.S.A.). An RNA-seq library was prepared with the KAPA Stranded mRNA-Seq kit following the manufacture's protocol in preparation for Illumina® sequencing. cDNA products within the 200-400 bp size range (obtained from mRNA fragmentation and cDNA synthesis) were selected by solid phase reversible immobilization using Agencourt AMPure XP reagent (Beckman Coulter, Inc. CA, U.S.A); this reagent was also used for all DNA cleanup steps within the procedure. PCR library amplification consisted of 14 PCR cycles. The cDNA library was checked for proper fragment size selection and quality using an Agilent 2100 Bioanalyzer. Library concentration was determined following the manufacture's protocol for the KAPA Library Quantification Kit (Illumina® platforms; KAPA Biosystems, Boston, MA, U.S.A.). The *P. sulphureus* venom gland cDNA library was then pooled with equal concentrations of seven other uniquely barcoded cDNA libraries, and sequenced on one lane of an Illumina® HiSeq 2000 platform at the UC Denver Genomics core to obtain 100 bp paired-end reads.

Transcriptome Assembly

The overall quality of sequenced reads were assessed using the Java program FastQC (Babraham Institute Bioninformatics, U.K.), and any low quality or adaptor sequences were identified and removed using the FASTX-Toolkit (Hannon lab, Cold Spring Harbor Laboratory). To obtain the best venom gland transcriptome assembly, two assembly approaches were used in combination. A Trinity (release v2014-07-17) *de novo* assembly of paired-end reads was completed using default parameters (Grabherr et al. 2011). A second *de novo*

assembly was completed with the program Extender (Rokyta et al. 2012a). For Extender, the reads were first merged with PEAR (Paired-end read mergeR; v0.9.6 using default parameters), if the 3' ends overlapped, to create longer contiguous sequences (Zhang et al. 2014). The top 1,000 merged reads were then used as seeds for the generation of complete transcripts. The same Extender parameters were set as used previously for other rear-fanged snake venom gland assemblies (McGivern et al. 2014); these were using 1,000 merged reads as seeds and extension of seeds required an overlap of 100 nucleotides, phred scores of at least 30 at each position in the extending read, and an exact match in the overlapping region. Contigs from these two assemblies were combined and BLASTx (executed using BLAST+ command line) was performed against a custom snake protein database using a minimum E-value of 10^{-4} (Camacho et al. 2009). The custom database was assembled from snake proteins within the NCBI (National Center for Biotechnology Information) database (accessed August 2015). Keywords searched included "venom" with "three-finger toxin", "metalloproteinase", "C-type lectin", "serine protease/proteinase", "phospholipase A_2", "L-amino acid oxidase", "cysteine-rich secretory protein", "ohanin", "phosphodiesterase", "kunitz-type protease inhibitor", "hyaluronidase", "nerve growth factor", "vascular endothelial growth factor", "vespryn", "cystatin", "venom factor", "bradykinin-potentiating peptide", "C-type natriuretic peptide", "nucleotidase", "acetylcholinesterase", "ficolin", "glutaminyl-peptide cyclotransferase", "exendin", "myotoxin", "waprin", "sarafotoxin", and "waglerin". All annotated transcripts (toxin and non-toxin) from assembled venom

gland transcriptomes of *Ophiophagus hannah* (Vonk et al. 2013)*, Micrurus fulvius* (Margres et al. 2013), *Crotalus adamanteus* (Rokyta et al. 2012b), *Crotalus horridus* (Rokyta et al. 2013), and *Boiga irregularis* and *Hypsiglena torquata* (McGivern et al. 2014), were also added to the reference set, as well as annotated transcripts (toxin and non-toxin) from the published genomes of *O. hannah* and *Python bivittatus* (Castoe et al. 2013; Vonk et al. 2013). To identify complete coding sequences, the resulting BLASTx output and all contigs were used as input files for the stand-alone ORFpredictor program (Min et al. 2005). The resulting predicted CDS (coding-sequence) and protein sequences from all contigs were then clustered with CD-HIT to remove any redundancy from multiple assemblies (Li and Godzik 2006; Fu et al. 2012). Reads were then aligned with Bowtie2 to predicted CDS, and transcript abundances were determined using RSEM (RNA-seq by Expectation-Maximization; v1.2.23) (Li and Dewey 2011). Transcripts below a FPKM abundance value of 1 were excluded from the analysis. The remaining transcripts were identified as venom proteins after another BLASTx search against the custom venom protein database and each manually examined to determine if the resulting venom protein produced was full-length, shared sequence similarities to currently known venom protein, and contained a shared signal peptide sequence with other venom proteins within that superfamily. This final set of venom protein transcripts was used as reference sequences for downstream proteomic analysis.

Whole Venom Trypsin Digest and Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

Approximately 100 µg of crude *P. sulphureus* venom and 20-35 µg of each of the three purified 3FTxs from *P. sulphureus* venom were subjected to LC-MS/MS analysis at Florida State University College of Medicine Translational Science Laboratory (Tallahassee, FL, U.S.A), performed using an LTQ Orbitrap Velos equipped with a Nanospray Flex ion source and interfaced to an Easy nanoLC II HPLC (Thermo Scientific). Crude venom and purified proteins were digested using the Calbiochem ProteoExtract All-in-one Trypsin Digestion kit (Merck, Darmstadt, Germany) with LC/MS grade solvents according to the manufacturer's instructions. The LC-MS/MS analyses were performed using an LTQ Orbitrap Velos equipped with a Nanospray Flex ion source and interfaced to an Easy nanoLC II HPLC (Thermo Scientific). Peptide fragments were separated using a vented column configuration consisting of a 0.1 x 20 mm, 3 µm C18 trap column and a 0.075 x 100 mm, 3 µm C18 analytical column (SC001 and SC200 Easy Column respectively, Thermo Scientific). The elution gradient consisted of 5% buffer B (0.1% formic acid in HPLC grade acetonitrile) and 95% buffer A (0.1% formic acid) at the run start, to 35% B at 60 min, to 98% B from 63-78 min with a flow rate of 600 nl/min from 64-78 min, and 5% B at 300 nl/min at 79 min. The mass spectrometer was operated in positive mode nanoelectrospray with a spray voltage of +2300 V. A "Top 9" method was used with precursor ion scans in the Orbitrap at 60K resolving power and fragment ion scans in the linear ion trap. Precursor ion selection using MIPS was enabled for charge states of 2+, 3+

and 4+. Dynamic exclusion was applied for 60 sec at 10 ppm. ITMS scans were performed using collision-induced dissociation (CID) at 35% normalized collision energy. MS/MS peptide spectra produced were interpreted using Mascot (Matrix Science, London, UK; version 1.4.0.288), Sequest (Thermo Fisher Scientific, San Jose, CA, U.S.A; version 1.4.0.288), and X! Tandem (thegpm.org; version CYCLONE 2010.12.01.1), assuming a trypsin digestion. The Mascot5_Trembl_bony vertebrate database, and the Sequest and X! Tandem Uniprot Serpentes (A8570) databases, were used for homology searches. Sequest and X! Tandem were searched with a fragment ion mass tolerance set to 0.6 Da and a parent ion tolerance of 10 ppm. Mascot was searched with a fragment ion mass tolerance of 0.8 Da and a parent ion tolerance of 10 ppm. Glu→pyro-Glu of the N-terminus, ammonia loss of the N-terminus, Gln→pyro-Glu of the N-terminus, carbamidomethylation of cysteines and carboxymethylation of cysteines were specified as variable post-translational modifications within X! Tandem. Oxidations of methionine, carbamidomethyl cysteine, and carboxymethyl cysteine were specified as variable post-translational modifications within Mascot and Sequest. Results were viewed and validated within Scaffold (Proteome Software Inc., Portland, OR, U.S.A; version 4.4.6), and protein identities were accepted if they could be established at >99.9% probability and contained at least one identified peptide. The normalized spectral abundance factor (NSAF) approach was used to quantify the abundance of each venom protein superfamily (Paoletti et al. 2006; Zybailov et al. 2006). This approach was used for both spectra identified using available databases and for

the spectra identified using the custom *P. sulphureus* venom gland toxin transcriptome reference.

Results and Discussion

Venom Yields and Venom Gland

Venom yields from *P. sulphureus* ranged from 250-500 µL, resulting in 3 to 9 mg of crude venom when lyophilized (n = 3). The venom gland of adult *Pseustes* is large, and like most rear-fanged colubrids, it is located posterior and ventral to the eye, is attached at the posterior and medial surface of the gland to a ligament running to the jaw rictus and lacks musculature associated with the gland (Fig. 7).

Pseustes sulphureus Crude Venom Enzyme Activity and Toxicity

No detectable enzyme activity was observed for L-amino acid oxidase, acetylcholinesterase, phosphodiesterase, phospholipase A_2, thrombin-like or kallikrein-like serine proteinase assays using *P. sulphureus* crude venom. Metalloproteinase activity was detectable, but relativity low at 0.150 (± 0.014) $\Delta A_{342\ nm}$/min/ per mg venom protein. Human fibrinogen, consisting of three subunits (α, β, and γ), was incubated with crude venom for an hour at 37 °C with no observable hydrolysis of any of the three subunits (data not shown). These enzymatic assays demonstrate that *P. sulphureus* crude venom largely lacks the enzymatic activity that is observed in many other snake venoms (Mackessy 2010b).

Despite lacking enzymatic activity, crude *P. sulphureus* venom was toxic towards both House Geckos and NSA mice (Table 5). LD_{50} values for both NSA mice and House Geckos were below 3 µg/g, with crude venom considerably more toxic towards geckos than mice. The murine LD_{50} value for *P. sulphureus* venom was similar to values reported for fairly toxic rear-fanged venomous snakes, such as *Alsophis portoricensis* (2.1 µg/g), but crude *A. portoricensis* venom was not nearly as toxic towards lizards as was *P. sulphureus* venom (Weldon and Mackessy 2010). When compared to the House Gecko LD_{50} value for crude venom from *Boiga irregularis* (2.5 µg/g), a snake with a known lizard prey-specific toxin, crude venom from *P. sulphureus* had an even lower LD_{50} (1.01 µg/g).

Table 5. Lethal toxicity (LD_{50}) of *Pseustes sulphureus* venom and purified sulmotoxins.

	Mice (*Mus musculus*)	House Geckos (*Hemidactylus frenatus*)
Crude Venom (*Pseustes sulphureus*)	2.56 (± 0.05) µg/g	1.01 (± 0.23) µg/g
Sulmotoxin A	Nontoxic >5 µg/g	0.22 (± 0.31) µg/g
Sulmotoxin B	3 – 5 µg/g	> 5 µg/g
Sulmotoxin C	Nontoxic > 5 µg/g	Nontoxic > 5 µg/g

Pseustes sulphureus Crude Venom Protein Profile

Crude venom from *P. sulphureus* when subjected to reducing one dimensional SDS-PAGE displayed at least eight separate protein bands (Fig. 8A). There were clusters of protein bands between 50-100 kDa and one band at approximately 25 kDa within the *P. sulphureus* crude venom profile, but the largest and most distinctive cluster of bands was observed from 6-9 kDa. Bands in this 6-9 kDa range were also observed for other rear-fanged snake species, including *Trimorphodon biscutatus lambda*, *Oxybelis fulgidus, Psammophis schokari, Rhamphiophis oxyrhynchus,* and all species of *Boiga* (Fig. 8A). Within venoms of rear-fanged snakes, including *Boiga, Oxybelis* and *Trimorphodon*, proteins found in the molecular mass range of 6-9 kDa are most commonly 3FTxs (Fry et al. 2003c; Mackessy et al. 2006). The 1D SDS-PAGE venom profile for *P. sulphureus* suggests that 3FTxs are highly abundant in this venom, similar to other rear-fanged snake venoms that are known to contain 3FTxs.

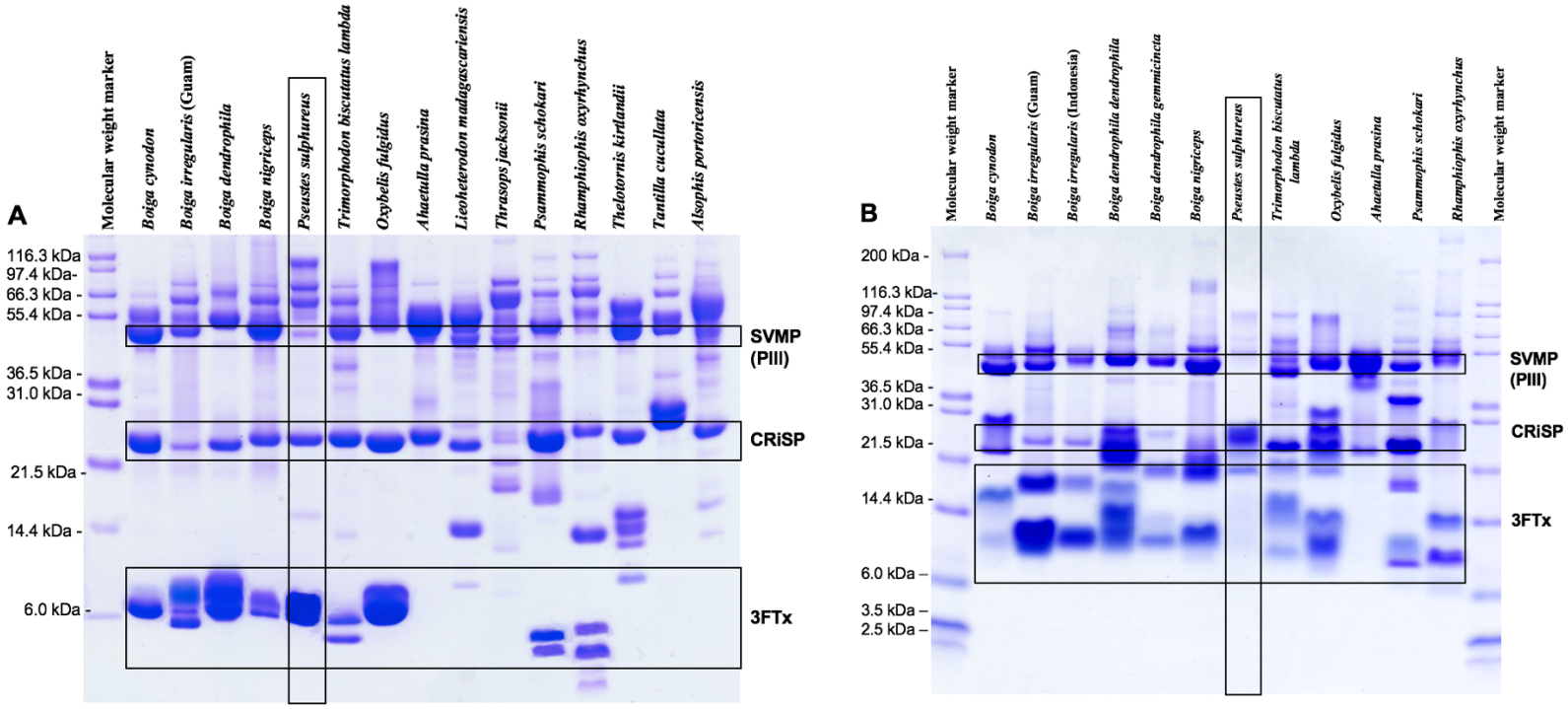


Figure 8. Reduced (A) and non-reduced (B) SDS-PAGE comparison of *Pseustes sulphureus* venom to other rear-fanged snake venom profiles. Molecular masses are indicated on the left, and putative venom protein identities are listed on the right. Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, and SVMP (PIII) = snake venom metalloproteinase type PIII.

Based on the molecular masses of proteins from other rear-fanged snakes, it is probable that the 25 kDa protein is a CRiSP (Peichoto et al. 2012). A transcript containing an open reading frame for a CRiSP was identified within the *P. sulphureus* venom gland transcriptome. This is also the same mass that was observed for a CRiSP present in the rear-fanged snakes *T. b. lambda* (Peichoto et al. 2012) and *B. irregularis* (Mackessy et al. 2006), also present on the gel (Fig. 8). Metalloproteinases (PIII) have been documented in rear-fanged snake venoms at a mass range of 49-53 kDa and are likely within *P. sulphureus* venom, which shows a band in this mass range and the presence of low SVMP enzyme activity within this venom.

Under non-reduced conditions for 1D SDS-PAGE, the lower protein band associated with 3FTxs shifted upward to approximately 17-19 kDa for *P. sulphureus* crude venom (Fig. 8B). This change in molecular mass for 3FTxs between reduced and non-reduced conditions has previously been observed for the 3FTx complex irditoxin within the venom of *Boiga irregularis* (Guam) (Pawlak et al. 2009). Irditoxin under reduced conditions displayed 3FTxs at approximately 7.5-8.5 kDa and under non-reduced conditions at approximately 17 kDa (Pawlak et al. 2009). Irditoxin was identified as a covalently linked heterodimeric three-finger complex and the stabilization of this covalent bond connecting two 3FTxs under non-reduced conditions resulted in the higher molecular mass (Pawlak et al. 2009). Given the similar observed approximate molecular masses changes between reduced and non-reduced one-dimensional SDS-PAGE for *P. sulphureus* crude venom, it is likely that *P. sulphureus* venom also contains an

abundant covalently linked 3FTx complex. This is only the second 3FTx complex documented within a rear-fanged snake venom.

Crude venom from *P. sulphureus* was also separated using reduced 2D SDS-PAGE for a more detailed venom profile. The most abundant venom proteins were observed at approximately 8-10 kDa with basic isoelectric points (Fig. 9), again suggestive of a venom largely dominated by 3FTxs. There were also slightly acidic venom proteins present at higher molecular masses (50-97 kDa) and more basic proteins at approximately 25 kDa, these protein masses were also observed in the one-dimensional SDS-PAGE crude venom profile.

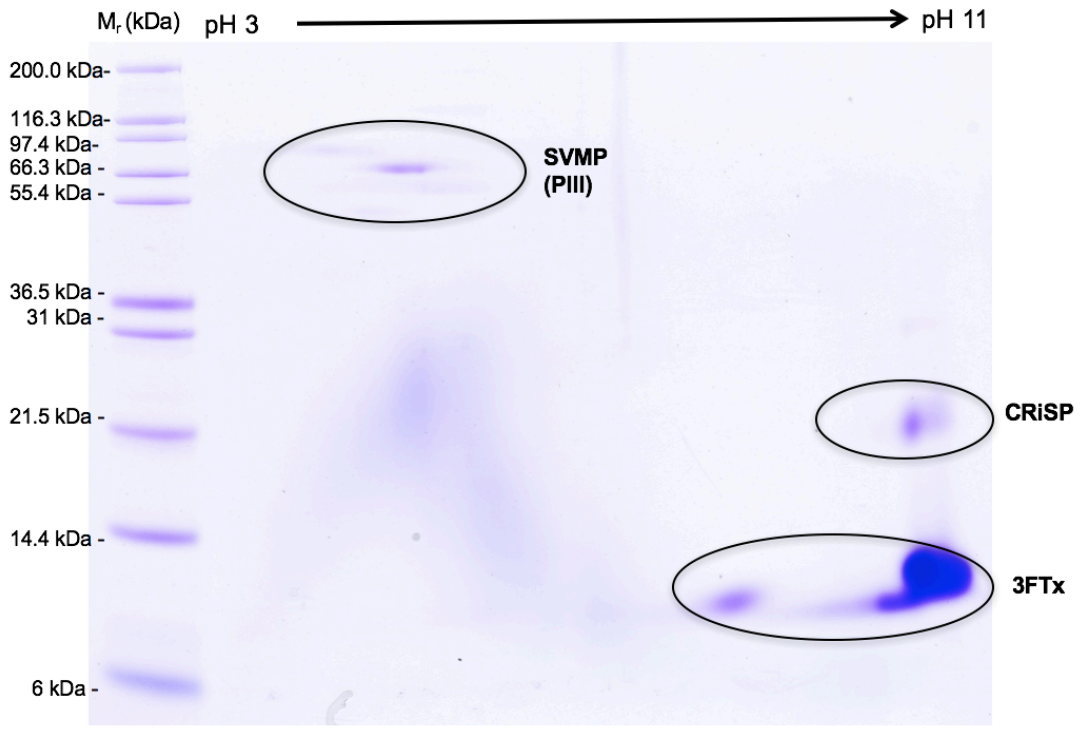


Figure 9. 2D gel electrophoresis *Pseustes sulphureus* crude venom profile. Crude venom (200 µg) from *P. sulphureus* was separated using an IPG pH 3-11 strip, reduced SDS-PAGE, and stained with Coomassie Brilliant Blue. Encircled areas highlight the presence of putative venom proteins. Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, and SVMP (PIII) = snake venom metalloproteinase type PIII.

To complement the *P. sulphureus* crude venom SDS-PAGE profiles, a MALDI-TOF crude venom profile was also generated. The more accurate MALDI-TOF masses still demonstrated an abundance of low molecular mass proteins, especially from 7,500 to 8,700 kDa (Fig. 10A). There was also an abundant protein at 17,259.22 Da, similar in mass to the 3FTx heterodimeric irditoxin (17,036.31 Da). Other prominent masses were seen at 24,980.57 Da (a CRiSP) and 55,423.01 Da (an SVMP PIII) (Fig. 10B).

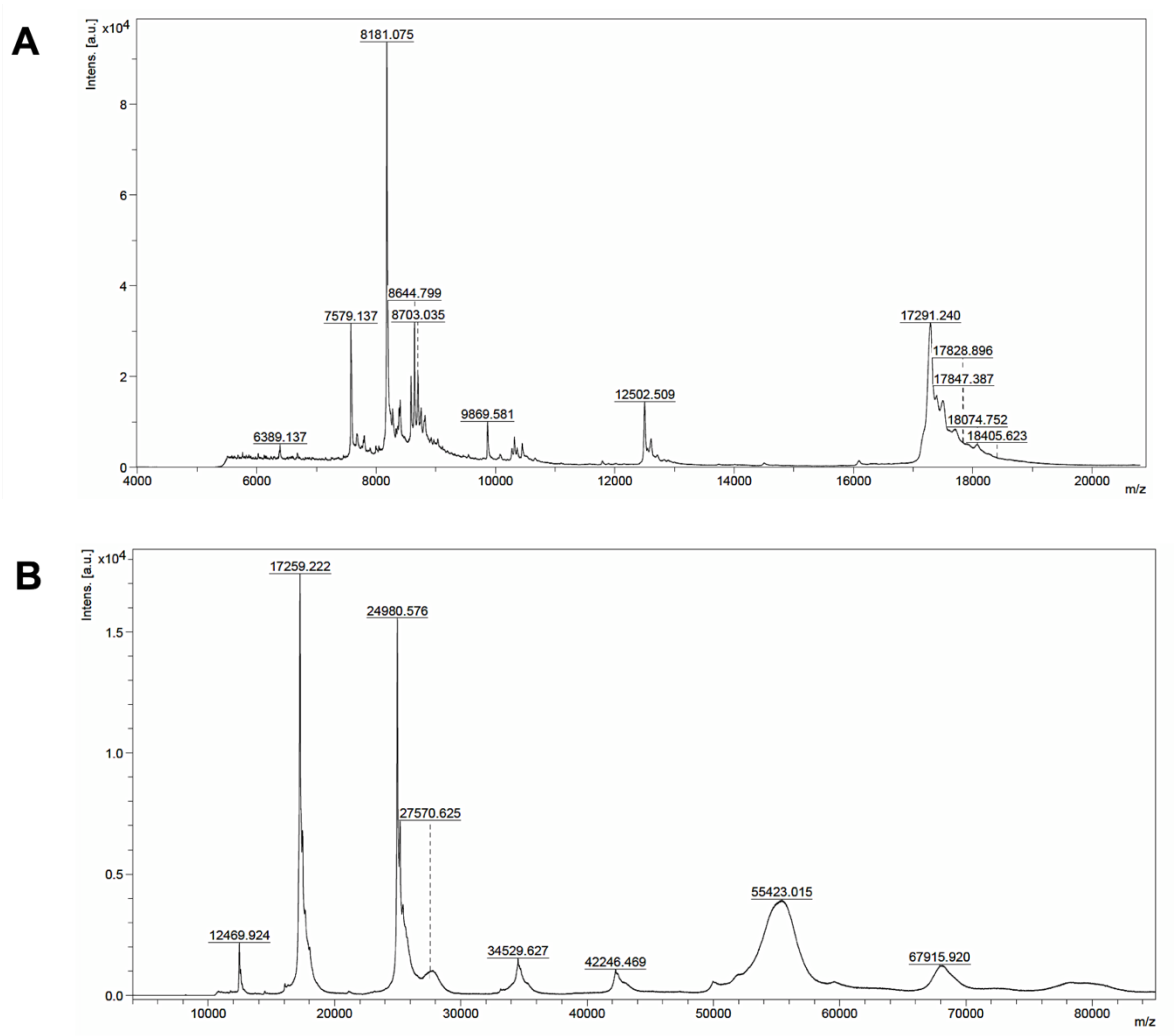


Figure 10. MALDI-TOF molecular masses of *Pseustes sulphureus* venom proteins. MALDI-TOF results from crude venom (4 µg) revealed a high abundance of smaller proteins with similar masses from 7-9 kDa (A) and also several prominent higher mass proteins (B) within the venom.

Size exclusion chromatography was used to profile crude *P. sulphureus* venom and as a first step in purifying venom proteins for identification and biological assays, and five to six peaks were resolved using size exclusion chromatography (Fig. 11A). Size exclusion fractionations of different individuals revealed intraspecific differences, with one venom displaying six peaks (Fig. 11A) and venoms from two other individuals displaying only five peaks (data not

shown). Size exclusion profile for *B. irregularis* venom is shown for comparison (Fig. 11B). However, regardless of the individual *Pseustes*, SDS-PAGE profiles for each peak were nearly identical (Fig. 11C).

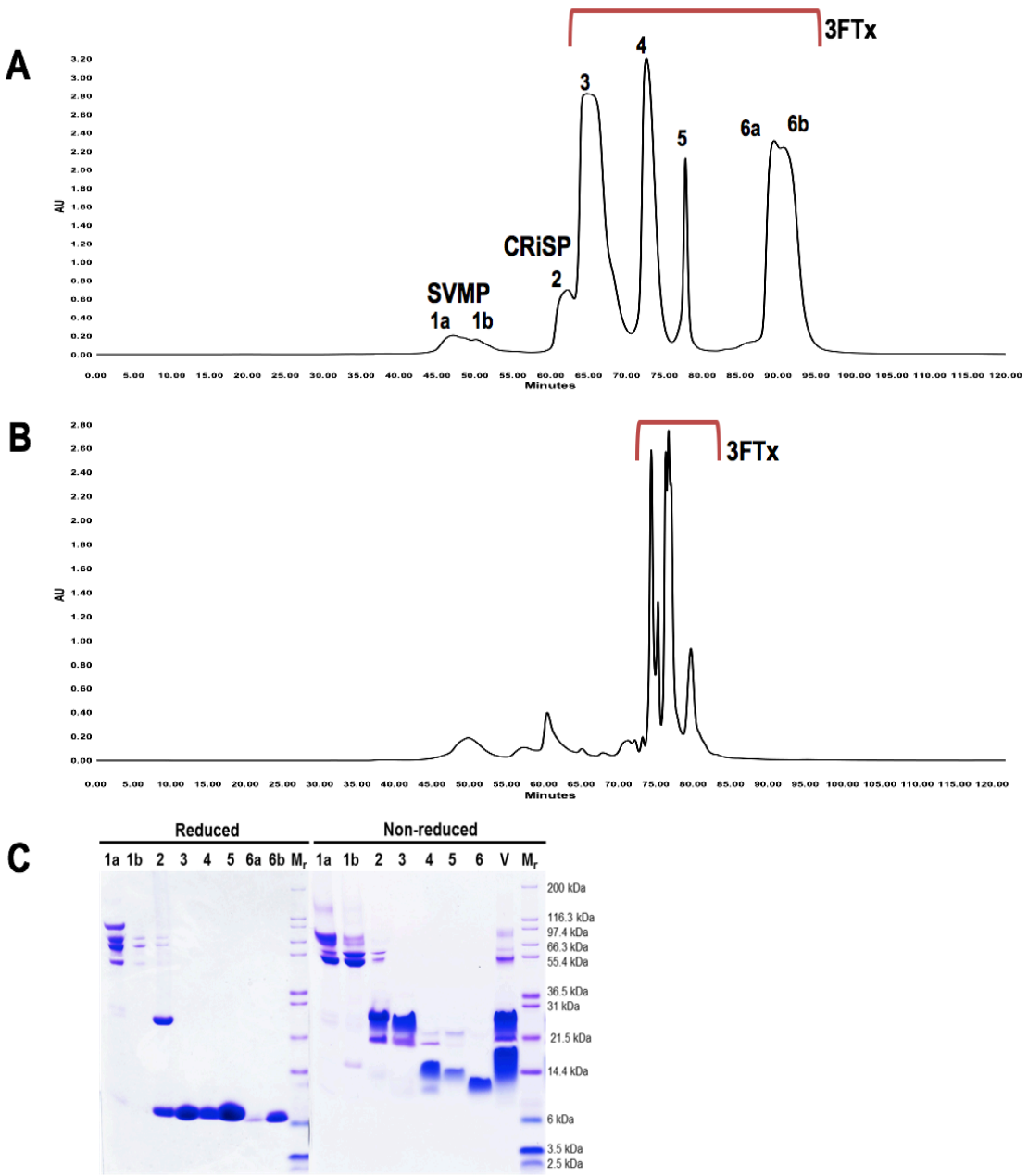


Figure 11. Size exclusion venom profiles for *Pseustes sulphureus* and *Boiga irregularis*, with SDS-PAGE results of *P. sulphureus* peaks. Crude venoms (3 mg) from *P. sulphureus* (A) and *B. irregularis* (B) were fractionated using size exclusion HPLC on a Yarra 3µm column; protein identities are based on SDS-PAGE peak results, N-terminal sequencing, LC-MS/MS, and enzymatic assays. Reduced and non-reduced SDS-PAGE gels are shown for the first *P. sulphureus* chromatogram (C); crude venom (V) was also run under non-reduced conditions. Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, and SVMP = snake venom metalloproteinase.

The most abundant venom proteins were in peaks 2 through 6. The highest peaks (3-6) comprised approximately 91% of the total peak area and only contained proteins of 6-8 kDa when reduced (Fig. 11C). Larger proteins were minimally abundant in the venom; peak 1 consisted of approximate 4% and peak 2 approximately 5% of the total peak areas. Non-reduced SDS-PAGE of peak 3 resulted in a protein band at approximately 17-18 kDa, similar to the protein mass change observed when crude *P. sulphureus* venom was subjected to reduced and non-reduced SDS-PAGE conditions (Fig. 11C). Results presented below showed that peak 3 corresponded with the heterodimeric 3FTx complex.

Venom from *Boiga irregularis,* known to be heavily dominated by 3FTxs (Pawlak et al. 2009; McGivern et al. 2014), was also fractionated under identical conditions and compared to the *P. sulphureus* venom profile (Fig. 11B). The overall size exclusion venom profile of *Boiga irregularis* showed a large cluster of abundant proteins eluting after 70 minutes. This abundant cluster of peaks are 3FTxs, based on the later elution time and the 6-8 kDa molecular mass from reduced SDS-PAGE (data not shown). The size exclusion venom profile for *P. sulphureus* shares many similarities with the *Boiga irregularis* venom profile, such as the high abundance of smaller proteins (3FTxs) and low abundance of larger venom proteins (CRiSPs, enzymes).

Low azocasein activity (0.115 ± 0.001 $\Delta A_{342\ nm}$/min/ per mg venom) was detected in *P. sulphureus* peak 1, and all other peaks lacked enzymatic activity. Azocaseinolytic activity is due to metalloproteinase(s) present within the first peak for *P. sulphureus*. Peak 2 was subjected to RP-HPLC (data not shown) to

obtain the purified protein product around ~22-23 kDa for N-terminal sequencing and protein identification. The purified ~22-23 kDa protein had an N-terminal sequence of DVDFDSESPRRQDKQKEIVD. This peptide sequence exhibited high similarity (95%) to a CRiSP found in *Telescopus dhara* (P84807.1). This sequence was also identical to the CRiSP venom protein transcript identified within the *P. sulphureus* venom gland transcriptome (see below). Peaks 3, 4, and 6 (Fig. 11A) were also subjected to RP-HPLC and produced a single purified protein product (Fig. 12A-C). Peak 5 was also within the 3FTx mass range, but lacked enough protein material for full characterization of this 3FTx. Peaks 3, 4, and 6 (sulmotoxins A, B, and C, respectively) were digested with trypsin and subjected to LC-MS/MS analysis using an Easy nanoLC II HPLC interfaced to a LTQ Orbitrap Velos with a Nanospray Flex ion source.

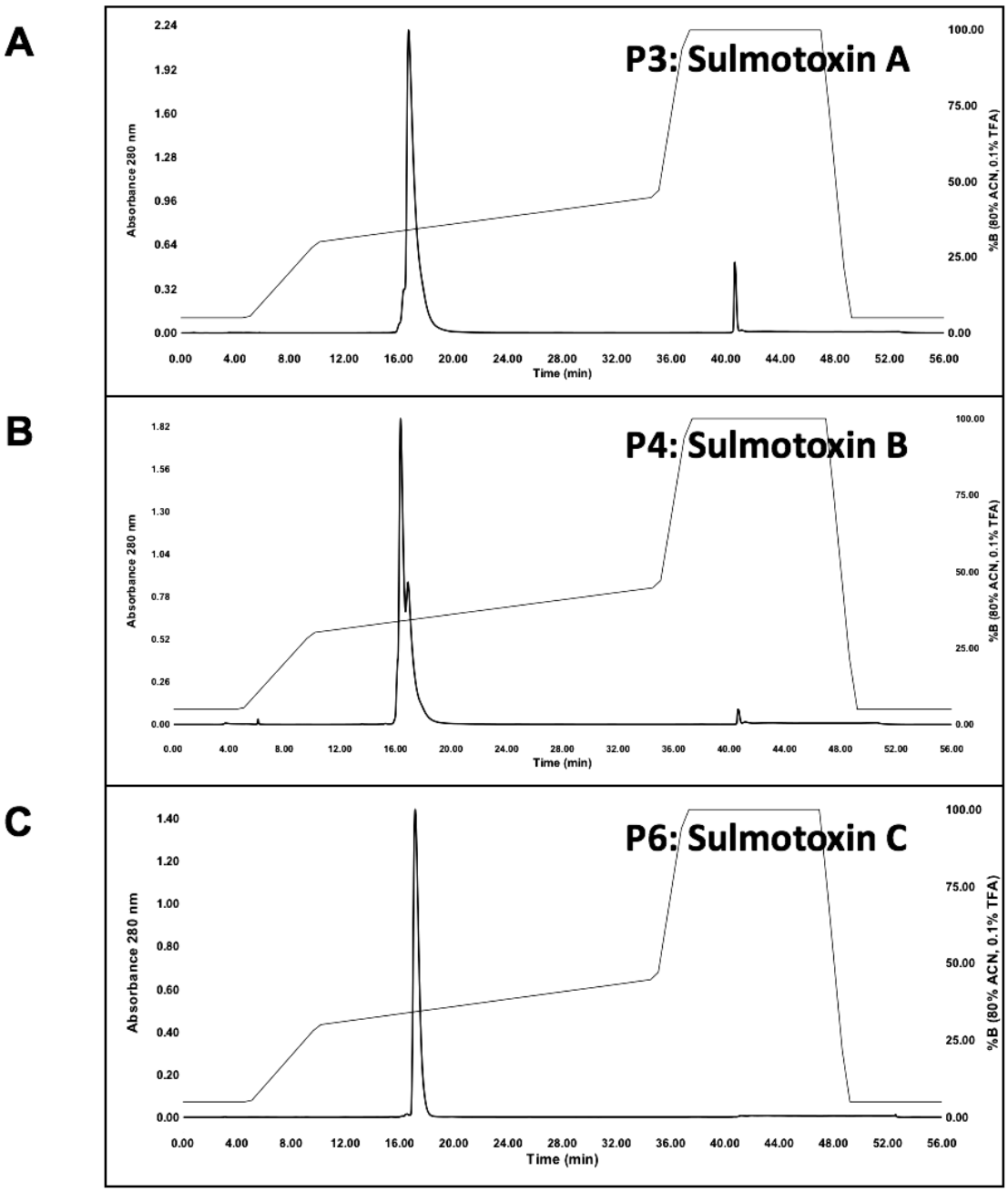


Figure 12. The three primary three-finger toxins (sulmotoxins A, B, and C) within *Pseustes sulphureus* venom were RP-HPLC purified (A-C). Sulmotoxins A, B, and C were RP-HPLC purified using a 30-45% linear gradient of acetonitrile (80% in 0.1% TFA) over 25 minutes. Sulmotoxin A (A) and sulmotoxin C (C) eluted as single peaks. Sulmotoxin B (B) showed a small shoulder peak, and the largest peak was used for identification and biological assays.

Mass spectra for peak 3, the 3FTx complex (sulmotoxin A), could not be identified using Mascot, Sequest, and X! Tandem databases. There are currently relatively few venom proteins from rear-fanged venomous snakes represented in these databases, and this novel 3FTx complex shared such a low sequence identity to the current 3FTxs present within these databases that it could not be identified with high probability (>99.9%) as a 3FTx. However, when the toxin transcripts from the *P. sulphureus* venom gland transcriptome were used as a reference to attempt identification of this protein, 418 peptide spectra matched with high probability (>99.9%) to 3FTx transcripts within the *P. sulphureus* venom gland transcriptome.

Similar results were obtained for the purified proteins from peak 4 (sulmotoxin B) and peak 6 (sulmotoxin C). For sulmotoxin B, only five peptide spectra matching a 3FTx within the current databases (three-finger toxin 1f from *Boiga irregularis*) were observed. Using the *P. sulphureus* venom gland toxin transcriptome reference, 616 peptide spectra from sulmotoxin B matched with high probability (>99.9%) to 3FTx transcripts. Sulmotoxin C had 66 peptide spectra that also matched a 3FTx within the current databases (three-finger toxin 1f from *Boiga irregularis*), but the number of identifiable spectra increased to 620 with the use of the *P. sulphureus* venom gland toxin transcriptome reference.

To determine which 3FTxs were the major toxic components of *P. sulphureus* venom, purified sulmotoxin A, sulmotoxin B, and sulmotoxin C were used for LD_{50} assays. A purified protein product for each 3FTx was obtained following RP-HPLC (Fig. 12A-C), with no other proteins detected with MALDI-

TOF MS. Sulmotoxin B had an isoform eluting as a small shoulder just after the main peak, but only the prominent peak was used. Once purified, accurate masses for each toxin were obtained with MALDI-TOF MS. Sulmotoxin A had a mass of 17,800 Da, sulmotoxin B a mass of 8,179.1, and sulmotoxin C a mass of 7,579.9 (Fig. 13).

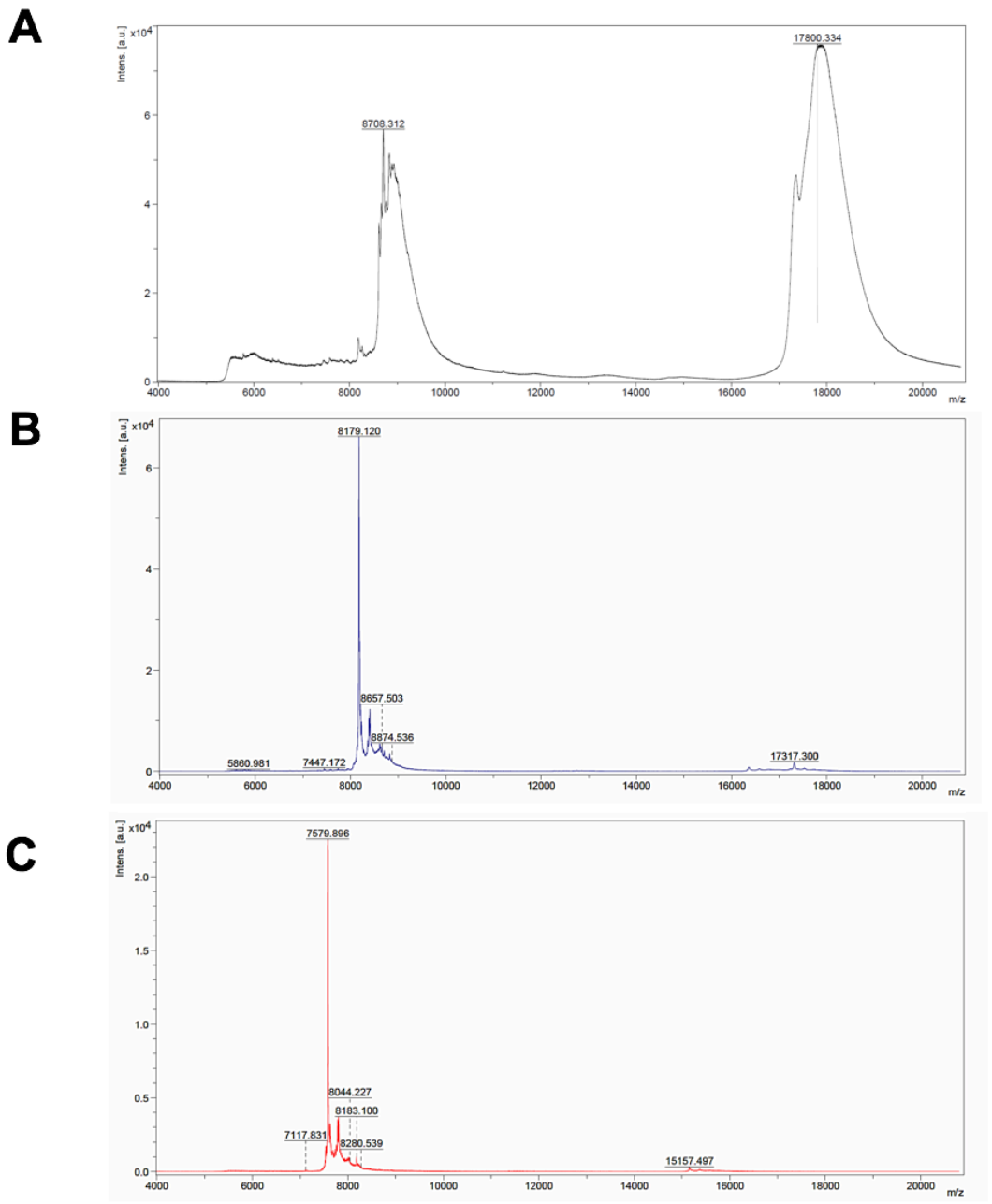
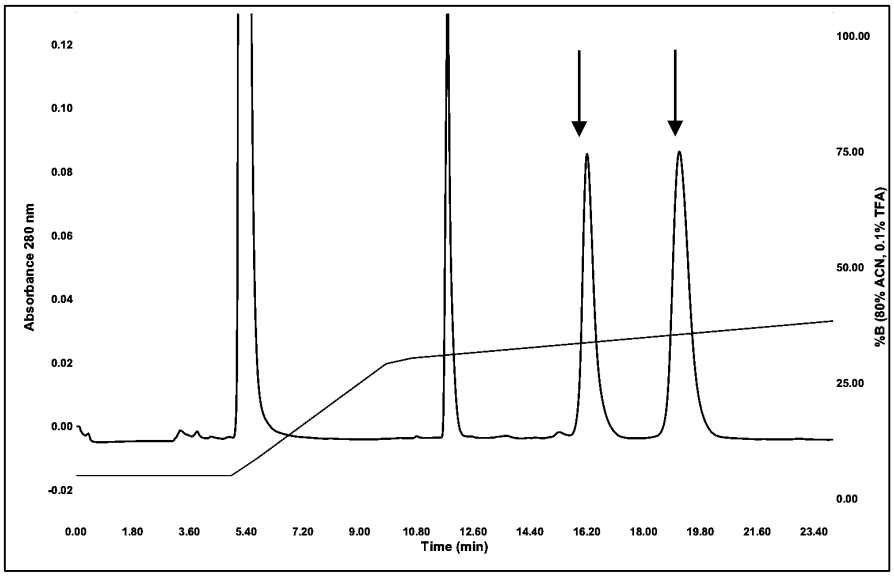


Figure 13. Sulmotoxin A, B, and C MALDI-TOF spectra. The molecular masses for the three most abundant 3FTxs, sulmotoxin A (A), sulmotoxin B (B), and sulmotoxin C (C), were determined.

Sulmotoxins A and B were reduced and alkylated and then subjected to RP-HPLC to evaluate potential subunit composition (Fig. 14A-B). Sulmotoxin A showed two peaks for the reduced and alkylated protein product and sulmotoxin B only one, demonstrating that sulmotoxin is a heterodimeric 3FTx complex and sulmotoxin B is a monomer. Reduced and alkylated sulmotoxin subunits A1 and A2 had masses of 9,338.7 Da and 9,221.3 Da, respectively (Fig. 15). Sulmotoxin A is only the second heterodimeric 3FTx complex that has been described from a rear-fanged snake venom.

A



B

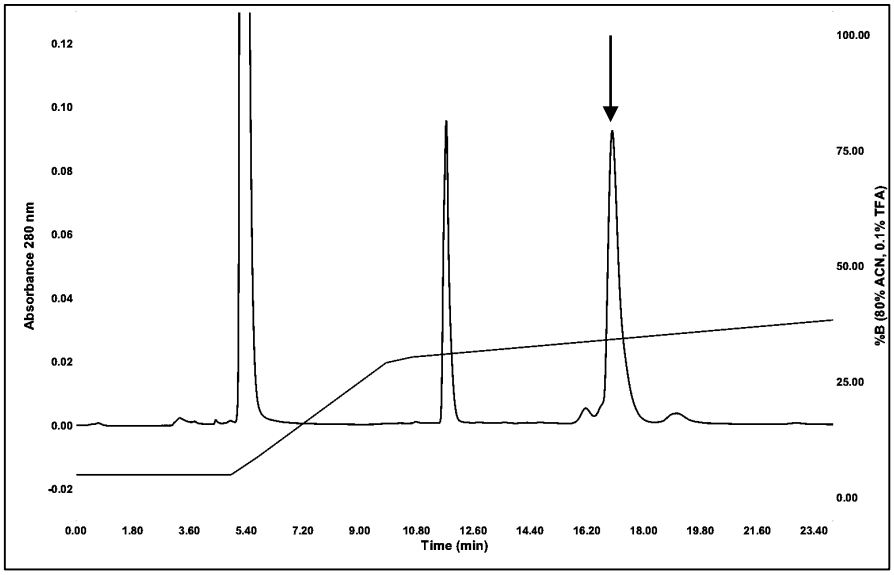


Figure 14. Sulmotoxin A (A) and sulmotoxin B (B) were reduced, alkylated, and subjected to RP-HPLC. The protein peaks are indicated with arrows; other peaks are from reduction/alkylation chemistry (DTT and IAA). Sulmotoxin A shows two protein peaks (two separate 3FTxs that had been covalently linked prior to reduction) and sulmotoxin B contains only one.

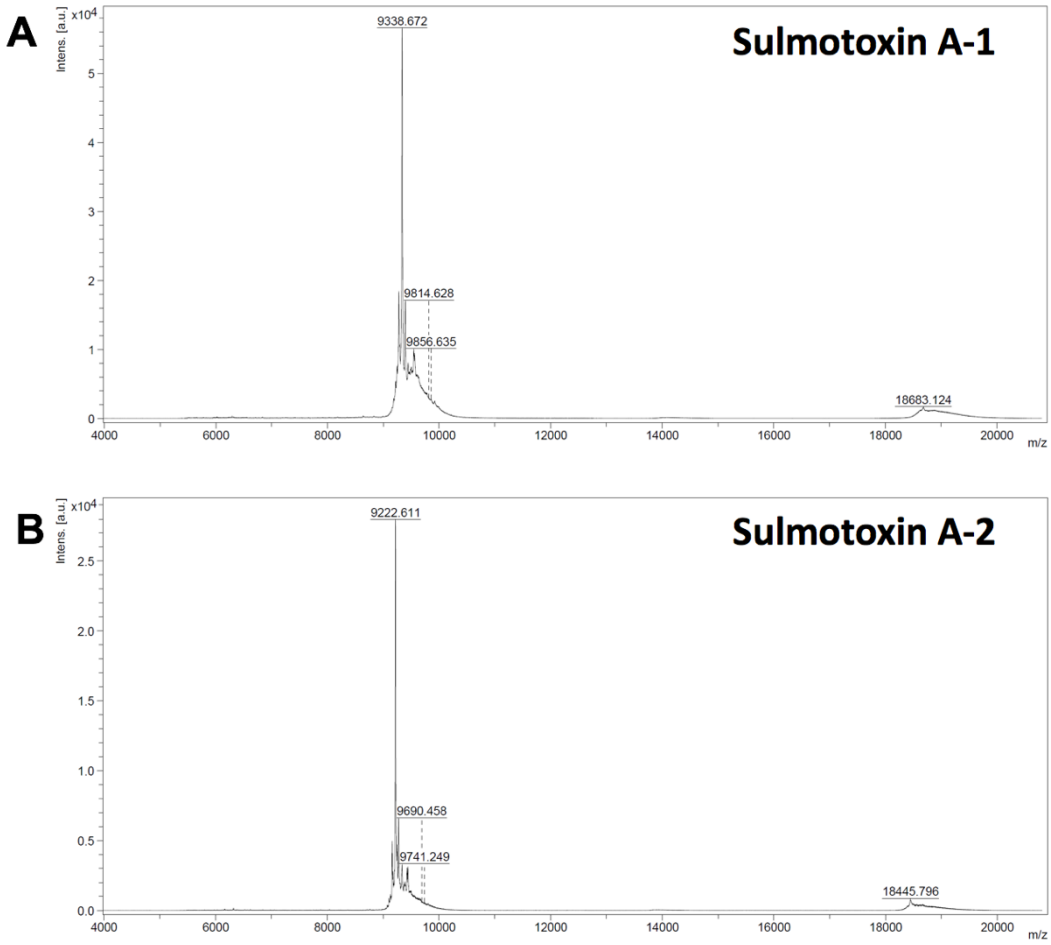


Figure 15. Masses of sulmotoxin A subunits A1 and A2 as determined by MALDI-TOF MS following reduction, alkylation and RP-HPLC purification.

Lethal toxicity (LD_{50}) assays were performed with sulmotoxins A, B, and C using House Geckos (*H. frenatus*) and NSA mice (*M. musculus*) (Table 5). Sulmotoxin A was found to have a LD_{50} value of 0.22 (± 0.31) µg/g in House Geckos (*H. frenatus*) and was non-toxic in mammals, with an LD_{50} value of >> 5 µg/g for mice (*M. musculus*). Sulmotoxin B exhibited the opposite trend, with a lower LD_{50} value for mice (*M. musculus*), approx. 4 µg/g, and non-toxic towards House Geckos (*H. frenatus*) (> 5 µg/g) (Table 5). Sulmotoxin C was non-toxic (>

5 µg/g) towards both House Geckos (*H. frenatus*) and NSA mice (*M. musculus*). These results demonstrate that the two most abundant 3FTxs within *P. sulphureus* venom display prey-specific toxicity, with sulmotoxin A exhibiting lizard-specific toxicity and sulmotoxin B showing mammal-specific toxicity. This is the first time that 3FTxs targeting different prey items in a single venom have been characterized and demonstrates the significance of 3FTxs as an important trophic adaptation.

Pseustes sulphureus Venom Gland Transcriptome Profile

A total of 29,462,311 paired-end reads (100 bps) were obtained from Illumina® HiSeq 2000 sequencing. Of these, 27,611,682 high-quality (phred scores greater than 30) paired-end reads were used for the *P. sulphureus* venom gland transcriptome assemblies. The Trinity *de novo* assembly generated 250,810 contigs with an average contig length of 1,133 bases and a N50 of 2,481 bases. Merging of overlapping pair-end reads resulted in 54% of reads merged, and these merged reads were used as input for the Extender assembly. Extender assembled 373 contigs. After ORFpredictor was used to obtain predicted CDS from all contigs and redundant sequences from the multiple assemblies were removed with CD-HIT-EST, a final total of 195,521 non-redundant contigs were assembled. Of these, 10,942 were expressed transcripts with a FPKM (fragments per kilobase of transcript per million mapped reads) value of >1, and 38 were full-length (complete CDS) venom protein transcripts that shared sequence similarity to other venom proteins and included a complete signal peptide sequence.

Of the 10,942 expressed transcripts within the *P. sulphureus* venom gland transcriptome, the large majority of the most highly expressed transcripts (FPKM > 8,000) were identified as toxin transcripts (Fig. 16A). There were 13 identified toxin superfamilies observed within the *P. sulphureus* venom gland transcriptome, including three-finger toxins (3FTxs), C-type lectins (CTLs), cysteine-rich secretory proteins (CRiSPs), L-amino acid oxidases (LAAOs), ficolins, snake venom metalloproteinases PIII (SVMPs PIII), venom factors (VF), phospholipase Bs (PLBs), 5’ nucleases (NUCs), venom endothelial growth factors (VEGFs), Kunitz-type protease inhibitors (KUNs), and phospholipase and metalloproteinase inhibitors (Fig. 16B and Table 6). Other venom protein families, such as serine proteinases, natriuretic peptides, acetylcholinesterases, and additional snake venom metalloproteinase PIII isoforms, were identified from partial transcripts. However, because these sequences do not contain a complete open reading frame, they were excluded from the venom gland toxin transcriptome.

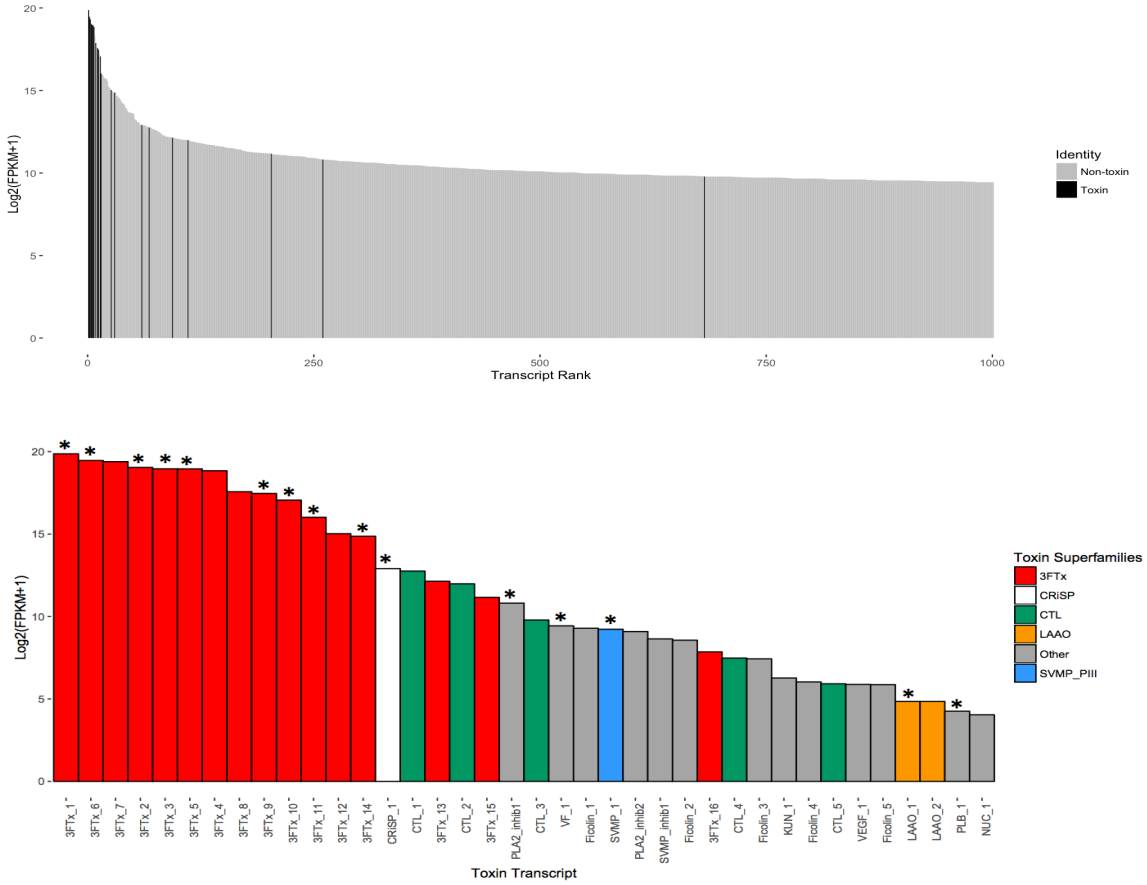


Figure 16. High levels of toxin transcripts are expressed within the *Pseustes sulphureus* venom gland, with three-finger toxins exhibiting the highest levels of expression. Toxin transcripts for translated venom proteins detected within the venom are marked with an asterisk. Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, CTL = C-type lectin, KUN = Kunitz-type protease inhibitor, LAAO = L-amino acid oxidase, PLA2_inhib = phospholipase A_2 inhibitor, PLB = phospholipase B, SVMP = snake venom metalloproteinase (PIII), SVMP_inhib = snake venom metalloproteinase inhibitor, NUC = 5' nucleotidase, VEGF = venom endothelial growth factor, and VF = venom factor.

Table 6. List of all full-length (complete CDS) venom protein transcripts from the *Pseustes sulphureus* venom gland. Transcripts assembled by Extender are marked with an asterisk; all other transcripts were assembled using Trinity.

Rank	Full-length Toxin Transcript ID	TPM (Transcripts Per Million)	FPKM (Fragments Per Kilobase of transcript per Million mapped reads)	Percentage of Toxin Reads
1	3FTx_1	96729	478480	4.53%
2	3FTx_6*	73058	361387	4.43%
3	3FTx_7*	69649	344528	4.42%
4	3FTx_2*	54551	269842	4.34%
5	3FTx_3*	51513	254814	4.32%
6	3FTx_5*	51372	254118	4.32%
7	3FTx_4*	47509	235010	4.29%
8	3FTx_8*	19657	97234	4.00%
9	3FTx_9	18299	90518	3.98%
10	3FTx_10	13831	68418	3.89%
11	3FTx_11*	6717	33227	3.65%
12	3FTx_12	3366	16650	3.42%
13	3FTx_14*	3028	14981	3.39%
14	CRiSP_1	776	3840	2.94%
15	CTL_1	699	3460	2.91%
16	3FTx_13	455	2252	2.76%
17	CTL_2	409	2025	2.73%
18	3FTx_15	231	1145	2.54%
19	PLA2_inhib1	181	899	2.46%
20	CTL_3	89	442	2.23%
21	VF_1	69	345	2.15%
22	Ficolin_1	63	312	2.12%
23	SVMP_1	60	299	2.10%
24	PLA2_inhib2	55	272	2.07%
25	SVMP_inhib1	40	199	1.97%
26	Ficolin_2	38	189	1.70%
27	3FTx_16	23	115	1.69%
28	CTL_4	18	89	1.43%
29	Ficolin_3	17	86	1.37%
30	KUN_1	7	38	1.35%
31	Ficolin_4	6	32	1.34%
32	CTL_5	6	30	1.34%

Table 6 continued.

33	VEGF_1	5	29	1.34%
34	Ficolin_5	5	29	1.34%
35	LAAO_1	2	14	1.10%
36	LAAO_2	2	14	1.10%
37	PLB_1	1	9	0.97%
38	NUC_1	1	8	0.92%

Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, CTL = C-type lectin, KUN = Kunitz-type protease inhibitor, LAAO = L-amino acid oxidase, PLA2_inhib = phospholipase A_2 inhibitor, PLB = phospholipase B, SVMP = snake venom metalloproteinase (PIII), SVMP_inhib = snake venom metalloproteinase inhibitor, NUC = 5' nucleotidase, VEGF = venom endothelial growth factor, and VF = venom factor.

The most abundant full-length open reading frames encoding venom protein superfamilies within the *P. sulphureus* venom gland corresponded to those encoding 3FTxs (60% of toxin reads), CTLs (11% of toxin reads), and ficolins (8.4% of toxin reads) (Table 6). The three venom protein superfamilies with the highest levels of toxin transcript expression also contained the greatest diversity of isoforms, with 16 3FTx isoforms and 5 isoforms each for CTLs and ficolins. Three-finger toxin transcripts were observed to be 20.69 – 91.86% identical, with all translated transcripts containing highly conserved cysteines presumed to form the five disulfide linkages characteristic of non-conventional 3FTxs (Fig. 17). Venom protein superfamily diversity for the *P. sulphureus* venom gland is comparable to that identified in other rear-fanged snake venom gland transcriptomes (McGivern et al. 2014).

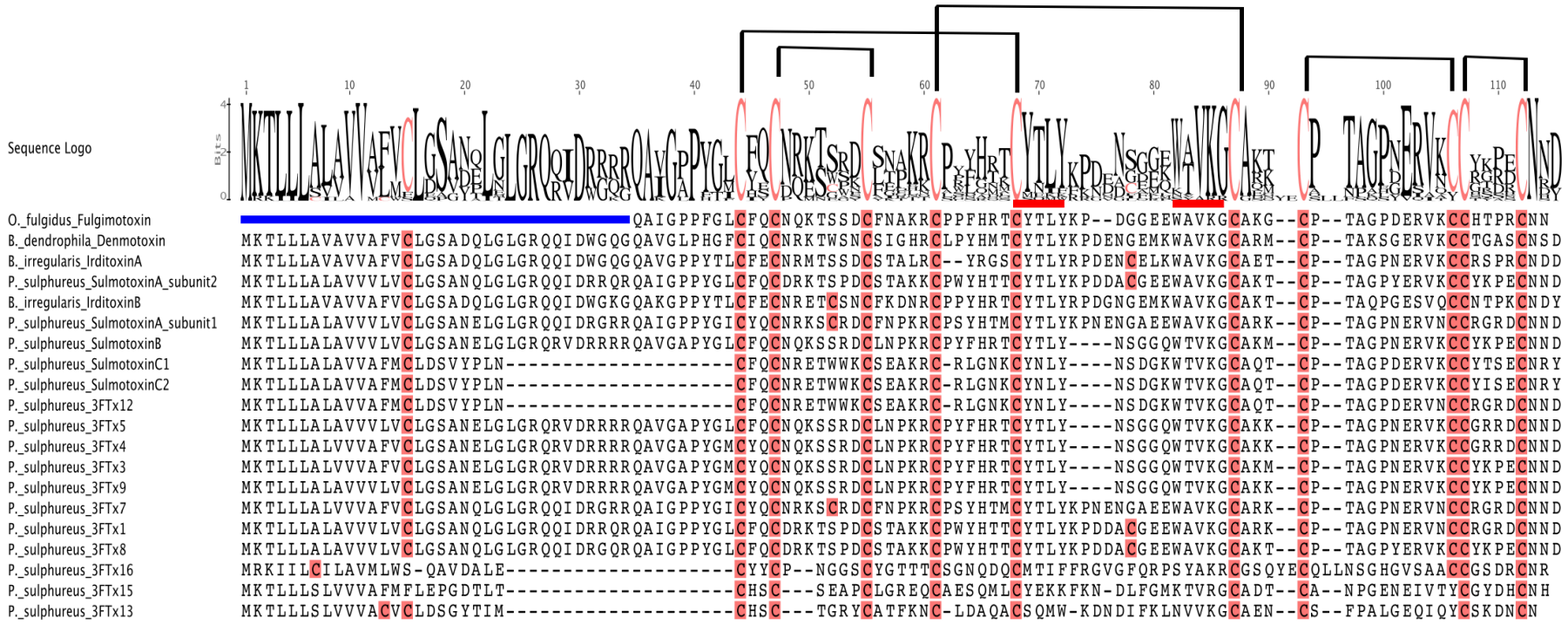


Figure 17. Sequence alignment of all *Pseustes sulphureus* three-finger toxin isoforms with known prey-specific non-conventional three-finger toxins from *Boiga dendrophila, Boiga irregularis* and *Oxybelis fulgidus*. Cysteines are highlighted; the conserved three-finger toxin structure results from constraints imposed by the disulfide linkages. The five disulfides characteristic of non-conventional 3FTxs are shown with solid lines. Rear-fanged snake venom prey-specific toxins denmotoxin (Q06ZW0.1), irditoxin subunit A (A0S864.1), irditoxin subunit B (A0S865.1), and fulgimotoxin (C0HJD3.1) are also shown in the alignment, with the putative prey-specific sequence motifs highlighted by red bars above the sequence. The signal peptides are indicated by the blue bar above all sequences except fulgimotoxin.

The total number of full-length venom protein transcripts (38) reported for *P. sulphureus* is less than that observed for the rear-fanged snakes *Boiga irregularis* (108 putative toxin transcripts) and *Hypsiglena* sp. (79) (McGivern et al. 2014), but much greater than the number reported for *Macropisthodon rudis* (5), a natricine rear-fanged snake (Zhang et al. 2015). The venom composition is relatively simple for *P. sulphureus* compared to the rear-fanged venomous snake species listed above, and this could be why *P. sulphureus* has less venom protein transcripts compared to *Boiga irregularis* and *Hypsiglena* sp. The *B. irregularis* venom gland transcriptome was described as elapid-like because of the dominant expression of three-finger toxins, and the rear-fanged venom gland transcriptome of *Hypsiglena* sp. was described as viperid-like because of the dominant expression of metalloproteinases (McGivern et al. 2014). Venom gene expression within the venom gland of *P. sulphureus* is more similar to the elapid-like venom gland transcriptome of *B. irregularis.* However, fewer three-finger toxin isoforms were observed within the *P. sulphureus* venom gland transcriptome than in the venom gland transcriptome of *B. irregularis.*

Extender was able to assemble abundant 3FTx isoforms better (Table 6) and assembled a total of 9/16 3FTx isoforms. The Trinity *de novo* assembly was better able to assemble full-length transcripts for low abundance toxins and toxin superfamilies with few isoforms present; however, use of Trinity alone for venom gland transcriptome assemblies has been reported to miss multi-copy venom genes/isoforms (Hargreaves and Mulley 2015; Macrander et al. 2015). In our

hands, combining the two assemblies allowed for better overall coverage of both multiple isoform protein families and low abundance toxin transcripts.

***Pseustes sulphureus*
Transcriptome-
Facilitated
Proteomics**

Using the LC-MS/MS spectra generated from each purified sulmotoxin in combination with the *P. sulphureus* venom gland toxin transcriptome reference and the MALDI-TOF masses of each sulmotoxin, transcripts corresponding to each of the sulmotoxins were identified. In a previous study (Heyborne and Mackessy, 2013) focused on prey-specific 3FTxs within rear-fanged snake venoms, a putative motif of prey-specific 3FTxs was proposed (Heyborne and Mackessy 2013). The amino acid sequences WAVK and CYTLY in the second 3FTx loop were identified as conserved motifs for all lizard/bird-specific 3FTxs and possible receptor binding site resulting in prey-specific toxicity. Sulmotoxin A also shares the WAVK and CYTLY sequences within the second 3FTx loop of both subunits. However, sulmotoxin B has the sequence WTVK, a substitution of alanine to threonine, which is a change from a nonpolar, hydrophobic amino acid to a polar amino acid. These results both support the possibility of the WAVK amino acid sequence being responsible for lizard-specific toxicity because sulmotoxin B, which lacks significant toxicity toward lizards, does not share this conserved amino acid sequence.

The high abundance of 3FTxs and relative paucity of larger enzymatic venom proteins within the *P. sulphureus* venom proteome reveals an elapid-like venom which parallels the high expression levels of 3FTx venom protein

transcripts within the venom gland transcriptome. Relatively simple venoms with high abundance of 3FTxs have been suggested to be characteristic of snake species with diets of a single vertebrate class, such as sea snakes feeding on teleost fish (Calvete et al. 2012). However, the identification of two different prey-specific 3FTxs within the venom of *P. sulphureus* indicates that snakes with relatively simple venoms can still target a diversity of prey types. Among the three finger protein superfamily, a wide variety of disparate pharmacologies has evolved on a highly conserved structural scaffold (Kini and Doley 2010), and within this one colubrid snake, the venome has capitalized upon this conserved yet versatile platform to produce a simple but effective trophic adaptation.

To obtain the identities of all proteins present within crude *P. sulphureus* venom, LC-MS/MS analysis was performed on trypsin digested crude venom. This also provided insight into the translation regulation of venom protein transcripts by comparing toxin transcript abundances (Fig. 18A) to the abundances of venom proteins (Fig. 18B-C). To identify the produced venom protein peptide spectra, generated spectra were searched using Mascot, Sequest, and X! Tandem against public databases, and protein identities were accepted if they could be established at >99.9% probability and contained at least one identified peptide. The percentage of each venom protein superfamily within *P. sulphureus* venom was determined using the normalized spectral abundance factor (NSAF) approach to quantify abundances (Paoletti et al. 2006; Zybailov et al. 2006) (Fig. 18B). This analysis was also completed using the *P. sulphureus* venom gland toxin transcriptome as a custom reference to identify

peptide spectra present within crude venom and to calculate protein abundances (Fig. 18C).

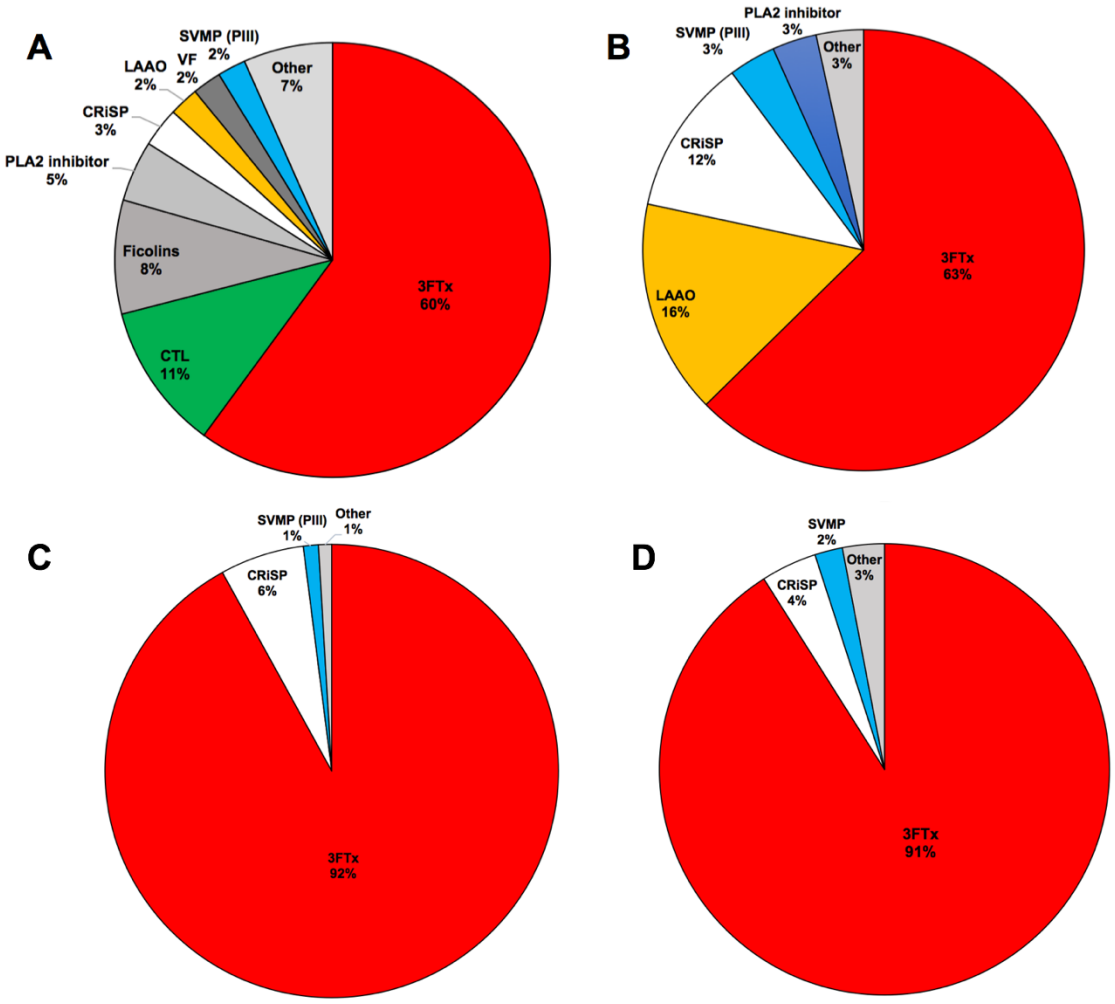


Figure 18. Venom gland toxin transcriptome and venom proteome abundance profiles for *Pseustes sulphureus*. Abundances for transcripts belonging to each venom protein superfamily are shown (A), and superfamilies making up less than 1% were combined to form the “other” category. The relative abundances of each superfamily within the venom are also shown; results using current databases as a reference for LC-MS/MS peptide identification (B) are compared with results using the *P. sulphureus* toxin transcriptome reference (C) and with protein abundances based on SE HPLC (D).

Using the current databases, a total of 9 venom protein families were identified within *P. sulphureus* venom. These included (in order of abundance) 3FTxs, LAAOs, CRiSPs, SVMPs (PIII), PLA_2 inhibitors, VEGFs, natriuretic peptides (NP), phospholipase A_2s, and serine proteases; the last four toxin families each represented less than <1% of all normalized venom protein spectra (Fig. 18B and Table 7). When using the *P. sulphureus* venom gland toxin transcriptome as a custom reference, 7 venom protein families were identified, and only 3FTxs, CRiSPs and SVMPs (PIII) each comprised greater than 1% of the total toxin normalized spectra (Fig. 18C and Table 8). Fifteen of the 38 toxin transcripts were identified as translated proteins within the venom (Fig. 16B).

Table 7. List of all MS/MS identified venom proteins from *Pseustes sulphureus* venom using Mascot, Sequest, and X! Tandem databases**.** All unique peptide fragments for venom proteins that had at least 10 total spectra are shown.

Protein Identity (Accession)	MS/MS-derived Peptide Sequences	Spectra	Protein superfamily
OXLA_CRODM	(K)HVVIVGAGMAGLSAAYVLAGAGHQVTVLEASER(V) (G)LSAAYVLAGAGHQVTVLEASER(V) (R)KFGLQLNEFFQENENAWYFIKNIR(K) (G)LQLNEFFQENENAWYFIK(N) (K)YILDKYDTYSTKEY(L) (K)EGNLSPGAVDMIGDLLNEDSGYYVSFIESLKHDDIFGYEKR(F) (K)RFDEIVGGMDQLPTSMYEAIKEK(V) (K)eKVQVHFNAR(V) (R)VIEIQQNDR(E) (R)RIKFEPPLPPKK(A) (R)IKFEPPLPPKKAHALR(S) (K)STTDLPSR(F) (R)WSLDKYAMGGITTFTPYQFQHFSEA(L) (R)WSLDKYAmGGITTFTPYQFQHFSEALTAPFKR(I) (T)PYQFQHFSEALTAPFKR(I) (R)IYFAGEYTAQFHGWIDSTIK(S) (R)IYFAGEYTAQFHGWIDSTIKSGLTAAR(D) (R)IYFAGEYTAQFHGWIDSTIKSGLTAARDVNRASENPSG(I) (Y)TAQFHGWIDSTIK(S)	122	LAAO
A0A0B8RSZ0_BOIIR	(K)GcAQTcPTAGPNERVK(C)	106	3FTx
T1E6V1_CROOH	(Y)VLAGAGHQVTVLEASER(V) (R)SGTKIFLTcTKK(F) (K)KFWEDDGIHGGKSTTDLPSR(F)	97	LAAO
A0A098LYK0_PANGU	(N)VDFDSESPR(R) (R)RSVSPTASNMLK(M) (R)SVSPTASNMLKMEWYSEAASNAER(W) (K)mEWYSEAASNAER(W) (K)MEWYSEAASNAERWAYR(C) (G)SVIGHYTQIVWYK(S)	34	CRiSP
tr|A0A0B8RS09|A0A0B8RS09_BOIIR	(K)EIVDMHNSFRR(S) (K)mEWYSEAASNAER(W) (N)PQGSVIGHYTQIVWYK(S) (K)QKEIVDMHNSFRR(S) (G)SVIGHYTQIVWYK(S)	33	CRiSP
A0A0B8RYY3_BOIIR	(K)GcAETcPTAGPNERVK(C)	17	3FTx
A0A0B8RSV9_BOIIR	(R)QENGVNIPcDPQDVK(C) (R)HcVNVQTAY(-)	13	SVMP PIII

Table 7 continued.

F8S106_CROAD	(K)LFLVADYIMYLK(Y) (R)MYDIVNVITPIYHR(M) (R)RKSHDNAQLLTGINFNGPTAGLAY LGGIcNTmYSAGIVQDHSK(I) (K)KPLKTDVVSPAVcGNYFVEVGEEc DcGSPR(T) (R)FKGAGTEcR(A) (R)AAKDEcDMADVcTGR(S) (K)DEcDmADVcTGR(S) (R)NGQPcKNNNGYcYNGK(C) (K)IAcEPQDVK(C) (K)cADGKAcSNGQcVDVTTPY(-)	12	SVMP PIII
tr|C0STK6|C0STK6_ELACL	(K)ADAQGAWHSAScDEDLLVVcEFSFI AFASVLER IPSPQLLNENKAFASVLER (E)NAFLTVHR(A)	11	PLA_2 inhibitor
PLIA_ELAQU	(K)cINIVGYR(K) (I)KGCVSSCPELR(L) (R)RNDLIKVEcR(D)	10	PLA_2 inhibitor

Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, PLA_2 = phospholipase A_2, and SVMP PIII = snake venom metalloproteinase type PIII, and LAAO = L-amino acid oxidase.

Table 8. List of all MS/MS identified venom proteins from *Pseustes sulphureus* venom using the venom gland toxin transcriptome as a reference. All unique peptide fragments for venom proteins that had at least 10 total spectra are shown.

Protein Identity (transcript)	MS/MS-derived Peptide Sequences	Spectra	Protein superfamily
3FTx_9	SSRDcLNPKR TcYTLYNSGGQWTVKGcAKK KcPTAGPNER VKccYKPEcNND	314	3FTx
3FTx_3	SSRDcLNPKR TcYTLYNSGGQWTVKGcAKK McPTAGPNER VKccYKPEcNND	292	3FTx
3FTx_11	cRLGNKcYNLYNSDGK LGNKcYNLYNSDGKWTVK cYNLYNSDGKWTVKGcAQTcPTA GPDER GcAQTcPTAGPDERVK WTVKGcAQTcPTAGPDERVK VKccYTSEcNRY	269	3FTx

Table 8 continued.

CRiSP1	AWTEIIQLWHDEDKHFIYGVGAN PQGSVIGHYTQIVWYK cVLNHSPDSSR cVLNHSPDSSRILEGIQcGENIYK EIVDMHNSFR HFIYGVGANPQGSVIGHYTQIVW YK ILEGIQcGENIYK MEWYSEAASNAERWAYR QKEIVDMHNSFR RSVSPTASNMLK SGPTCGDCPSAcDNGLcTNPcL HQNK SLVQQSNcQDDWIK SNcPAScFcHNEII SVSPTASNMLKMEWYSEAASN AER	189	CRiSP
3FTx_6	(K)ScRDcFNPKR(C) (K)RcPSYHTMcYTLYKPNENGA EEWAVKGcAR(K) (R)KcPTAGPNERVNccR(G)	109	3FTx
3FTx_1	cPWYHTTcYTLYKPDDACGEEW AVKGCARK (R)KcPTAGPNERVNccR(G)	87	3FTx
SVMP PIII_1	AcIMSATQGFEPSYQFSScSYEQ NLR cFKENQKINGVGFcR ETDLLANPDR FKGAGEVcR GAKDDcDFPEHcTGQSAEcPTDL FQR GLFSEDYSETHYSPDGR INGVGFcRQENGVNIPcDPQDVK ITSLVASVMAHELGHNLGIDHDT AScNcPAK KYTGNLTAIR LDIWSNKDK QENGVNIPcDPQDVK SNDNAQLLTGINFNGPTVGLAY MSNLcKPDHSVGIVQDHSK STIIVTPEQNKYLKAPK TRVYGIVNEVNVMFR VLNIHVALIR WRETDLLANPDR YLELFIVVDNVMYRK	82	SVMP PIII

Table 8 continued.

VF_1	(K)LILNTPLNIQSLPITVR(T) (R)GNANSLNQIQYFTYLILTK(G) (K)SDFGcTAGSGQNNLGVFEDAG LALATSTK(L) (R)SDFEDEFFGDDNIISR(S) (R)SDFPESWLWLTELLPEQPNSQ GISSK(T) (R)AVPFVIVPLQQGLHDIEVK(A) (R)NRWEEYNAR(T) (R)THNIEGTSYALLALLK(M) (K)FYHPDKATGLLNK(I) (R)cAEETcSLLNQQK(K)	33	VF
LAAO_1	FGLQLNEFFQENENAWYFIKNIR	30	LAAO
PLA2_inhib_1	GLTFLTVLNLNNNPIR LQTLLNWVQKDNVDLFSKEDVVc AFPK NLPHLHSLDLSGNLLEDLAPELFK NTGLTHVPAGLFQSLK SIAPNTFHGTPTLWILSLR	16	PLA_2 Inhibitor

Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, PLA_2 = phospholipase A_2, SVMP (PIII) = snake venom metalloproteinase type PIII, VF = venom factor, and LAAO = L-amino acid oxidase.

Although current databases identified a greater number of venom protein superfamilies than the custom database, this result is because only full-length (complete CDS) venom protein transcripts were used for the custom *P. sulphureus* venom gland toxin transcriptome reference. However, the overall number of spectra identified significantly using this transcriptome reference: without the reference, only 507 peptide spectra were matched to a venom protein, and with the reference transcriptome, 1,424 peptide spectra could be assigned. Because fewer spectra were excluded, the calculated abundances of each venom protein superfamily more accurately matched results from size exclusion chromatography of crude *P. sulphureus* venom (Fig. 18D).

One of the major differences observed between the pie diagrams is the difference in abundances of LAAO within the venom (Fig. 18B-C). As mentioned

previously, LAAO activity was not detected in any of the size exclusion peaks, and it was also not detected in the venom. It is possible that the LAAOs within *P. sulphureus* venom do not recognize l-kynurenine as a substrate. However, it is more likely that because LAAO enzymes are highly conserved across venoms, do not have many isoforms present per species, and are not rapidly evolving (Rodrigues et al. 2009; Aird et al. 2015), the *P. sulphureus* LAAO peptide spectra, in spite of being derived from a trace venom component, were matched with high probability to LAAO venom proteins within the public databases. This resulted in a potential bias in the number of spectra associated with the LAAO superfamily compared with any other superfamily. For example, using the public databases, *P. sulphureus* 3FTx peptide spectra are less likely to be assigned to a venom protein within the database as a result of rapid and diversifying evolution producing high isoform diversity and low sequence conservation between toxins from different species of venomous snakes (or even within a single venom) (Pawlak et al. 2006; Pawlak and Kini 2008; Doley et al. 2009; Sunagar et al. 2013). The exclusion of unmatched 3FTx spectra therefore produced inaccurate quantification of the relative abundance of 3FTxs and of the other venom protein superfamilies.

Conclusions

To characterize the venom composition of *P. sulphureus* (Amazon Puffing Snake), we applied a combination of enzymatic, transcriptomic, and proteomic methods. The resulting data provide important insights into rear-fanged snake venom composition. First, unlike other venomous snakes, *P. sulphureus* largely

lacks venom enzymatic activity and is dominated by 3FTxs. Within the exception of minor SVMP activity, this venom did not display any other detectable enzyme activity.

The venom gland transcriptome for *P. sulphureus* contained transcripts from 13 venom protein superfamilies, and the assembled transcriptome contains a diversity of venom protein superfamilies with full-length transcripts (complete CDS) equivalent to that observed for other rear-fanged venom gland transcriptomes (McGivern et al. 2014). This indicates that transcripts are stable for at least 12-24 hr following the death of the snake, even without typical precautions to preserve mRNA integrity.

The venom gland transcriptome was largely dominated by 16 3FTx isoforms (61% of toxin reads), with C-type lectin (11%), and ficolin (8.4%) superfamilies the second and third most abundant transcripts, respectively. However, C-type lectins and ficolins were not detected in the venom proteome, whereas the 3FTxs comprised greater than 90% of the total protein present. The venom transcriptome also contained 6 additional venom protein superfamilies: C-type lectins, ficolins, venom factosr, SVMP inhibitors, kunitz-type protease inhibitors, and 5'-nucleotidases. Fifteen of the 38 full-length (complete CDS) venom protein transcripts (39.5%) were detected as translated venom proteins within the crude venom.

The use of the *P. sulphureus* venom gland transcriptome proved to be essential as a reference for the identification of proteins within the venom proteome. Venom protein sequences currently in the Mascot, Sequest, and X!

Tandem databases lacked sufficient peptide sequence similarity to the two most abundant 3FTxs (sulmotoxin A and sulmotoxin B) within *P. sulphureus* venom to allow for identification of these toxins from known sequence databases. This highlights the fact that snake venoms that are not well characterized, such as the large majority of rear-fanged snake venoms, still are poorly represented by databases utilized for MS/MS spectra matching, resulting in poor coverage by shotgun proteomics methods. Our results illustrate the importance of a combined transcriptomic and proteomic to characterize currently unstudied snake venoms, and it is to be expected that similar poor coverage would result for other understudied proteomes.

Rear-fanged snake venoms typically do not have an abundance of enzymatic activities as is typical of viperid front-fanged venomous snakes but there are a few exceptions known (*A. portoricensis* and *Philodryas* sp.) (Acosta et al. 2003; Weldon and Mackessy 2010; Peichoto et al. 2012). Venom from *P. sulphureus* is one of the first rear-fanged snake venoms assayed which nearly lacks enzymatic activity, and only low metalloproteinase activity was detected in crude venom. The venom is dominated by non-enzymatic 3FTxs, comprising approximately 91-92% of the overall venom composition (as determined by both % peak areas from size exclusion HPLC and LC-MS/MS using toxin transcript references). This high percentage of 3FTxs is similar to observations of venoms from some front fanged elapid venomous snakes (Calvete et al. 2012), and it is noteworthy that this bias was observed in both the venom gland transcriptome and the venom proteome of *P. sulphureus*.

The 3FTxs present within *P. sulphureus* result in crude venom that is significantly toxic (< 3 µg/g) toward both House Geckos (*H. frenatus*) and NSA mice (*M. musculus*). The abundant toxic components of this venom are the two primary 3FTxs (sulmotoxins A and B). Sulmotoxin A is very toxic (0.22 µg/g) towards lizards and non-toxic (> 5 µg/g) towards mammals, while sulmotoxin B is moderately toxic towards mammals (approx. 4 µg/g) and non-toxic (> 5 µg/g) toward lizards. *Pseustes sulphureus* therefore possesses a venom with individual prey-specific proteins that facilitate handling of diverse prey types. This diversified activity is in contrast to observations for the front-fanged *N. kaouthia*, where only one 3FTx (α-cobratoxin) was lethal toward both lizards and mammals. Variation in venom compositional “strategies” among diverse species is becoming more commonly observed for both rear-fanged and front-fanged venomous snakes, and as noted above, a relatively simple venom with toxins dominated by one venom protein superfamily can still allow for the targeting of a diversity of prey.

Sulmotoxin A exhibits the conserved amino acid sequences in the second 3FTx loop, a feature shared with other lizard-specific 3FTxs from rear-fanged venomous snakes and absent from non-taxon-specific 3FTxs; sulmotoxin B lacks this motif and lacks lizard-specific toxicity. Therefore, the importance of this site to the evolution of prey-specific toxicity was supported by current results. Biological assays of purified 3FTxs, in combination with the complete transcriptome and venom proteome for *P. sulphureus*, have provided a comprehensive view into the evolution and diversification of trophic roles of

venom 3FTxs. In order to understand the multiple factors driving the evolution of venoms in squamate reptiles, it is critical to study the venoms of the most diverse clades of venomous snakes, the rear-fanged “colubrid” snakes. This study represents an important step toward realization of that goal.

CHAPTER IV

VENOM GENE EXPRESSION WITHIN REAR-FANGED VENOMOUS SNAKES AND THEIR CORRESPONDING VENOM PROTEOMES

Cassandra M. Modahl[a], Seth E. Frietze[a,b], and Stephen P. Mackessy[a*]
[a]School of Biological Sciences, University of Northern Colorado, 501 20th St., Greeley, CO 80639-0017, USA
[b]Present address:

Abstract

Background

Snakebite is a significant cause of human morbidity and mortality worldwide, and venoms of many front-fanged species (cobras, vipers, etc.) have been studied extensively. However, venoms of rear-fanged snakes are poorly known. The rear-fanged snake species that have been studied have been found to have venoms with many proteins quite similar to those from front-fanged species. The present work uses high-throughput technologies to compare venom gland gene expression and venom proteomes for two rear-fanged snakes, *Alsophis portoricensis* (Puerto Rican Racer) and *Ahaetulla prasina* (Asian Green Vine Snake). The venom from *A. portoricensis* has a high abundance of snake venom metalloproteinases (SVMPs), and one of the most abundant metalloproteinases, alsophinase, produced hemorrhage in mouse models; venom from *A. prasina* is currently uncharacterized. Characterizing the venom gland transcriptome, venom proteome, and venom biological activity of rear-

fanged species are critical steps necessary for ultimately unraveling the evolutionary trajectory of venom production among advanced squamate reptiles.

Results

Both venom gland transcriptomes were largely dominated by SVMPs (70% of toxin reads for *A. portoricensis* and 62% for *A. prasina*). However, the venom gland from *A. prasina* also contained many three-finger toxin (3FTx) transcript isoforms. *Ahaetulla prasina* 3FTx isoforms were more similar to those found in African rear-fanged snakes (*Dispholidus* and *Thrasops*) and Asian cobras (*Naja* and *Ophiophagus*) than to other Asian rear-fanged snakes. Conversely, the single *A. portoricensis* 3FTx transcript was more similar to Asian rear-fanged snake (*Boiga*) toxins, and like them, it possesses an extended N-terminus. Both venom proteomes consisted primarily of SVMPs (81% of the *A. portoricensis* proteome and 75% of the *A. prasina* proteome), and the venoms had high proteolytic activity, comparable to the high levels seen in many viperid snake venoms. Human fibrinogen subunits were rapidly degraded when incubated with either venom.

Conclusions

Although rear-fanged venomous snakes possess a less derived venom delivery system, both *A. prasina* and *A. portoricensis* express venom protein transcripts within their venom glands at levels similar to those observed for front-fanged venomous snakes. The venom gland transcriptomes and proteomes of *A. prasina* and *A. portoricensis* were dominated by metalloproteinases and their crude venoms exhibited high enzymatic activity, similar to what is seen in viperid

venomous snakes. These venoms potentially contribute to digestive biological roles within these snakes.

Introduction

Snake venoms contain a variety of proteins and peptides that function primarily in prey immobilization and digestion, and secondarily as a mechanism of defense (Mackessy 2010b). Venoms have allowed for advanced snakes (caenophidian snakes) to transition away from the use of constriction (commonly seen in basal henophidian snakes) and instead rely on a chemical means of prey capture (Kardong et al. 1997). Rear-fanged venomous snakes represent unique evolutionary lineages where a mechanically less complex venom delivery system has evolved as a trophic adaptation that in many ways parallels the more complex injection system of front-fanged snakes. Rear-fanged venomous snakes have a relatively low-pressure venom delivery system and lack the large venom storage reservoir (lumen) that is seen in front-fanged venomous snakes. Within the rear-fanged snake venom gland, venom proteins are produced and stored intracellularly and are released more slowly into a main duct conducting venom to the base of rear maxillary teeth, which may be grooved or modified, but never hollow (hollow fangs are only seen in front-fanged venomous snakes) (Kardong and Lavin-Murcio 1993).

Rear-fanged snake venoms have remained largely unexplored, and this dearth of knowledge is in strong contrast to the extensive research that exists on the venoms of front-fanged snakes. Venom research has focused primarily on venoms from the Elapidae (cobras, kraits, mambas, etc.) and Viperidae (vipers

and pit vipers), species that produce significantly larger venom yields and are responsible for the vast majority of human envenomations (Mackessy 2010b). Most rear-fanged venomous snakes are unable to deliver sufficient quantities of venom to produce systemic envenomation effects, but at least five species (*Dispholidus typus, Thelotornis capensis, Rhabdophis tigrinus, Philodryas olfersii*, and *Tachymenis peruviana*) are believed to have caused human fatalities (Kuch and Mebs 2002; Mackessy 2002; Prado-Franceschi and Hyslop 2002; Weinstein et al. 2011; Weinstein et al. 2013).

In general, rear-fanged snake venoms show lower complexity than the venoms of front-fanged snakes, commonly expressing only 20-40 protein spots (visualized by two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis; 2D SDS-PAGE), while front-fanged snake venoms show considerably higher complexity, commonly expressing venoms with well over 100 protein spots (Peichoto et al. 2012). Venom composition is likely closely linked to snake diet (Mackessy 1988; Daltry et al. 1996b; Andrade and Abe 1999; Li et al. 2005a; Pawlak et al. 2009), with rear-fanged snake venoms currently containing the only known examples of prey-specific toxins (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013). In several cases, rear-fanged snake venoms have been documented to contain novel protein superfamilies or venom proteins with unique evolutionary trajectories (such as the evolution of prey-specific toxins within these venoms (Pawlak et al. 2009; OmPraba et al. 2010; Ching et al. 2012).

The majority of venom protein superfamilies have representatives in most snake venoms, including those from many rear-fanged species (Mackessy 2002). Some of the most prominent superfamilies include phospholipases A_2 (PLA_2s), serine proteinases, snake venom metalloproteinases (SVMPs), three-finger toxins (3FTxs), proteinase inhibitors, and lectins (Mackessy 2010b). Myotoxic metalloproteinases have been isolated from the venoms of several rear-fanged snakes and these venoms have been reported to have proteolytic activity up to 25 times greater than some pit vipers (Peichoto et al. 2007; Weldon and Mackessy 2012; Sanchez et al. 2014). Metalloproteinases are also common components of viperid venoms and are responsible for local and systemic hemorrhage often seen following viper bites (Fox and Serrano 2010). One of these metalloproteinases, alsophinase, was characterized from the venom of *Alsophis portoricensis* (Puerto Rican Racer), a New World rear-fanged snake (Weldon and Mackessy 2012).

Alsophis {*Borikenophis*} *portoricensis*, is a rear-fanged “colubrid” snake (family Dipsadidae) native to numerous islands in the Caribbean. They are ground-dwelling, diurnal snakes with a diet consisting primarily of lizards (*Anolis* sp.) and *Eleutherodactylus* frogs (Rodriguez-Robles and Leal 1993; Weldon and Mackessy 2010). *Ahaetulla prasina* (family Colubridae; Asian Green Vinesnake), is native to large areas of Southeast Asia. It is an arboreal snake with a diet of small nestling birds, lizards, and frogs (Lim and Lim 1992). These snakes both have similar dietary preferences, but occupy very different ecological niches, so they represent diverse model species in which to explore the importance of these

two (potentially important) factors affecting venom composition. The present work explores venom gene expression within *A. portoricensis* and *A. prasina* venom glands and compares gene expression to venom proteome composition. By characterizing the venom gland transcripts, the venom proteome, and the biological activity of proteins within the venoms, a better understanding of the adaptive significance of specific venom proteins to the snakes will be revealed, as well as identification of any potential human health hazards these snakes could pose (Kardong 2002; Weinstein et al. 2013).

Materials and Methods

Materials

TRIzol reagent was purchased from Life Technologies (San Diego, CA, U.S.A.). The KAPA Stranded mRNA-Seq kit and KAPA Library Quantification Kit (Illumina® platforms) were purchased from KAPA Biosystems (Boston, MA, U.S.A.). Agencourt AMPure XP reagent was from Beckman Coulter, Inc. (Brea, CA, U.S.A). Novex Mark 12 unstained molecular mass standards, MES running buffer, LDS sample buffer, and precast 12% Bis-Tris NuPAGE electrophoretic gels were obtained from Life Technologies (San Diego, CA, U.S.A.). Pierce BCA protein assay kit was obtained from Thermo Fisher Scientific (Rockford, IL, U.S.A). Phospholipase A_2 assay kit was purchased from Cayman Chemical Co (Ann Arbor, MI, U.S.A.). All other reagents (analytical grade or better), were obtained from Sigma-Aldrich (St. Louis, MO, U.S.A.). All reagents used for molecular work were certified nuclease-free.

Snake Venom and Venom Gland Tissue Collection

Ahaetulla prasina imported from Indonesia and a *Alsophis portoricensis* originating from the Guana Island, British Virgin Islands were maintained in the University of Northern Colorado Animal Resource Facility in accordance with UNC-IACUC protocol #9204. Both snakes were adults, with *A. prasina* measuring 1,000 mm snout-to-vent and weighing 150 grams, and the *A. portoricensis* measuring 580 mm snout-to-vent and weighing 75 grams. Venom was manually extracted from both snakes using the method of Hill and Mackessy (1997) with subcutaneous injections of ketamine-HCl (20-30 mg/kg) followed by pilocarpine-HCl (6 mg/kg) (Hill and Mackessy 1997). Venom was centrifuged at 10,000 rpm for 5 minutes, frozen at -80 °C, lyophilized, and stored at -20 °C until use. Four days post-extraction, when mRNA levels are highest (Rotenberg et al. 1971), both snakes were humanely euthanized and venom glands dissected from each snake. Tissue from each of the venom glands (right and left glands) from *A. prasina* was placed directly into the TRIzol reagent for immediate RNA isolation. Gland tissue from *A. portoricensis* was placed into RNAlater solution and stored at -80 °C for five years before it was used to generate a cDNA library. All procedures were approved by the UNC Institutional Animal Care and Use Committee (IACUC protocol 9204.1).

RNA Isolation, cDNA Library Preparation, and Next-Generation Sequencing

RNA isolation was completed following the manufacture's protocol for the TRIzol reagent with an overnight -20°C incubation in 300 µL 100% ethanol with 40 µL 3 M sodium acetate to increased RNA yields before total RNA was resuspended in 16 µL nuclease-free H_2O. Poly-A+ RNA was selected from 4 µg of total RNA with oligo-dT beads using the KAPA Stranded mRNA-Seq kit (KAPA Biosystems, Boston, MA, U.S.A.). cDNA library preparation was completed with the KAPA Stranded mRNA-Seq kit following the manufacture's protocol in preparation for Illumina® sequencing. cDNA products within the 200-400 bp size range (obtained from mRNA fragmentation and cDNA synthesis) were selected by solid phase reversible immobilization using Agencourt AMPure XP reagent (Beckman Coulter, Inc. CA, U.S.A), this reagent was also repeatedly use for all DNA cleanup steps within the procedure. PCR library amplification consisted of a total of 14 PCR cycles. The cDNA library was checked for proper fragment size selection and quality using an Agilent 2100 Bioanalyzer. Library concentration was determined following the manufacture's protocol for the KAPA Library Quantification Kit (Illumina® platforms; KAPA Biosystems, Boston, MA, U.S.A.). Each venom gland cDNA library was then pooled with equal concentrations of six other uniquely barcoded venom gland cDNA libraries, and sequenced on one lane of an Illumina® HiSeq 2000 platform at the UC Denver Genomics core to obtain 100 bp paired-end reads.

Transcriptome Assembly

The overall quality of sequenced reads were assessed using the java program FastQC (Babraham Institute Bioninformatics, U.K.), and any low quality or adaptor sequences were identified and then removed using the FASTX-Toolkit (Hannon lab Cold Spring Harbor Laboratory). To obtain the best venom gland transcriptome assembly, two assembly approaches were used in combination. A Trinity (release v2014-07-17) *de novo* assembly of paired-end reads was completed with default parameters (Grabherr et al. 2011). A second *de novo* assembly was completed with the program Extender (Rokyta et al. 2012a). For Extender, the reads were first merged with PEAR (Paired-End read mergeR; v0.9.6 using default parameters) if their 3' ends overlapped to create longer contiguous sequences (Zhang et al. 2014). The top 1,000 of these merged reads were then used as seeds for the generation of complete transcripts. The same Extender parameters were set as used previously for other rear-fanged snake venom gland assemblies (McGivern et al. 2014). Contigs from these two assemblies were combined and BLASTx (executed using BLAST+ command line) performed to search the contigs against a custom snake protein database with a minimum E-value of 10^{-4} (Camacho et al. 2009). This custom database was assembled from snake proteins within the NCBI (National Center for Biotechnology Information) database (accessed August 2015), keywords searched included "venom" with "three-finger toxin", "metalloproteinase", "C-type lectin", "serine protease/proteinase", "phospholipase A_2", "L-amino acid oxidase", "cysteine-rich secretory protein", "ohanin", "phosphodiesterase", "kunitz-type

protease inhibitor", "hyaluronidase", "nerve growth factor", "vascular endothelial growth factor", "vespryn", "cystatin", "venom factor", "bradykinin-potentiating peptide", "C-type natriuretic peptide", "nucleotidase", "acetylcholinesterase", "ficolin", "glutaminyl-peptide cyclotransferase", "exendin", "myotoxin", "waprin", "sarafotoxin", and "waglerin". In addition, all annotated transcripts (toxin and non-toxin) from assembled venom gland transcriptomes of *Ohiophagus hannah* (Vonk et al. 2013)*, Micrurus fulvius* (Margres et al. 2013), *Crotalus adamanteus* (Rokyta et al. 2012b), *Crotalus horridus* (Rokyta et al. 2013), *Boiga irregularis,* and *Hypsiglena torquata* (McGivern et al. 2014), were also added to the reference set, as well as annotated transcripts (toxin and non-toxin) from the published genomes of *O. hannah* and *Python bivittatus* (Castoe et al. 2013; Vonk et al. 2013). To identify complete coding sequences, the resulting BLASTx output and all contigs were used as input files for the standalone ORFpredictor program (Min et al. 2005). The resulting predicted CDS (coding-sequence) and protein sequences from all contigs were then clustered with CD-HIT to remove any redundancy from multiple assemblies (Li and Godzik 2006; Fu et al. 2012). Reads were then aligned with Bowtie 2 to predicted CDS and transcript abundances determined using RSEM (RNA-seq by Expectation-Maximization; v1.2.23) (Li and Dewey 2011). Transcripts below a FPKM abundance value of 1 were excluded from the analysis. The remaining transcripts were identified as venom proteins after another BLASTx search against the custom venom protein database and each manually examined to determine if the resulting produced venom protein was full-length, shared sequence identity to a currently known

venom protein, and contained a shared signal peptide sequence with other venom proteins within that superfamily.

Whole Venom Trypsin Digest and Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

Approximately 100 µg of crude venom from each snake was sent to Florida State University College of Medicine Translational Science Laboratory (Tallahassee, FL, U.S.A) for LC-MS/MS analysis, which was performed using an LTQ Orbitrap Velos equipped with a Nanospray Flex ion source and interfaced to an Easy nanoLC II HPLC (Thermo Scientific). Crude venoms were digested using the Calbiochem ProteoExtract All-in-one Trypsin Digestion kit (Merck, Darmstadt, Germany) with LC/MS grade solvents according to the manufacturer's instructions. The LC-MS/MS analyses were performed using an LTQ Orbitrap Velos equipped with a Nanospray Flex ion source and interfaced to an Easy nanoLC II HPLC (Thermo Scientific). Peptide fragments were separated using a vented column configuration consisting of a 0.1 x 20 mm, 3 µm C18 trap column and a 0.075 x 100 mm, 3 µm C18 analytical column (SC001 and SC200 Easy Column respectively, Thermo Scientific). The elution gradient consisted of 5% buffer B (0.1% formic acid in HPLC grade acetonitrile) and 95% buffer A (0.1% formic acid) at the run start, to 35% B at 60 min, to 98% B from 63-78 min with a flow rate of 600 nl/min from 64-78 min, and 5% B at 300 nl/min at 79 min. The mass spectrometer was operated in positive mode nanoelectrospray with a spray voltage of +2300 V. A "Top 9" method was used with precursor ion scans in the Orbitrap at 60K resolving power and fragment ion scans in the linear ion

trap. Precursor ion selection using MIPS was enabled for charge states of 2+, 3+ and 4+. Dynamic exclusion was applied for 60 sec at 10 ppm. ITMS scans were performed using collision-induced dissociation (CID) at 35% normalized collision energy. MS/MS peptide spectra produced were interpreted using Mascot (Matrix Science, London, UK; version 1.4.0.288), Sequest (Thermo Fisher Scientific, San Jose, CA, U.S.A; version 1.4.0.288), and X! Tandem (thegpm.org; version CYCLONE 2010.12.01.1), assuming a trypsin digestion. The Mascot5_Trembl_bony vertebrate database, and the Sequest and X! Tandem Uniprot Serpentes (A8570) databases, were used for homology searches. Sequest and X! Tandem were searched with a fragment ion mass tolerance set to 0.6 Da and a parent ion tolerance of 10 ppm. Mascot was searched with a fragment ion mass tolerance of 0.8 Da and a parent ion tolerance of 10 ppm. Glu→pyro-Glu of the N-terminus, ammonia loss of the N-terminus, Gln→pyro-Glu of the N-terminus, carbamidomethylation of cysteines and carboxymethylation of cysteines were specified as variable post-translational modifications within X! Tandem. Oxidations of methionine, carbamidomethyl cysteine, and carboxymethyl cysteine were specified as variable post-translational modifications within Mascot and Sequest. Results were viewed and validated within Scaffold (Proteome Software Inc., Portland, OR, U.S.A; version 4.4.6), and protein identities were accepted if they could be established at >99.9% probability and contained at least one identified peptide. The normalized spectral abundance factor (NSAF) approach was used to quantify the abundance of each venom protein superfamily (Paoletti et al. 2006; Zybailov et al. 2006). This

approach was used for both spectra identified using available databases and for the spectra identified using the custom rear-fanged venom gland toxin transcriptome reference.

Protein and Enzyme Assays

Pierce BCA protein assay kit (Thermo Fisher Scientific, Rockford, IL, U.S.A) was used to determine the protein concentration for crude venom and individual purified proteins for all assays described. Metalloproteinase activity was determined using azocasin as a substrate and 20 µg of crude venom (Aird and da Silva 1991). This activity was expressed as $\Delta A_{342\ nm}$/min/ per mg venom protein. L-amino acid oxidase activity was assayed according to Weisbach (1961) with 20 µg of crude venom, and the activity was expressed as nmole of product formed/min/mg protein (Weisbach et al. 1961). Acetylcholinesterase activity was determined using 15 µg crude venom, incubated with the acetylthiocholine iodide substrate in a cuvette at 37°C (Ellman et al. 1961). Absorbance at 412 nm was taken every 10 seconds for ten minutes and the linear portion of the graph was used to calculate activity at µmole product formed/minute/mg venom protein. Phosphodiesterase activity was assayed with 20 µg crude venom using 1 mM bis-p-nitrophenylphosphate as substrate, following the protocol developed by Laskowski (1980); activity was reported as $\Delta A_{400\ nm}$/min/mg protein. Phospholipase A_2 activity was determined using a commercially available kit (Cayman Chemical Co.) as described by the manufacturer, using 2 µg crude venom in 200 µL total volume. Absorbance was measured at 414 nm every minute for five minutes and activity was reported as

μmole product formed/minute/mg protein. Thrombin-like and kallikrein-like serine proteinase activity was assayed according to Mackessy (1993) with benzoyl-Phe-Val-Arg-paranitroaniline substrate to determine thrombin-like activity and benzoyl-Pro-Phe-Arg-paranitroaniline substrate to determine kallikrein-like activity (Mackessy 1993b). The thrombin-like and kallikrein-like substrates were incubated with 20 μg crude venom and activity recorded as nmol of product formed/minute/mg protein. All enzyme assays described above were performed in triplicate. Fibrinogenase activity was determined using 20 μg of crude venom incubated with human fibrinogen (final concentration 0.5 mg/mL) at 37°C in a total volume of 200μL for periods of 0, 1, 5, 10, 30, and 60 minutes (Ouyang and Huang 1979). Twenty μL of this reaction mixture was removed at each time point and mixed with an equal volume of 4%SDS and 5% 2-mercaptoethanol, then heated in boiling water for 10 minutes. Five μL aliquots were combined with 2x LDS buffer, electrophoresed on a 12% NuPAGE Bis-Tris gel, stained with Coomassie Brilliant Blue overnight, destained for two hours, and imaged.

Results and Discussion

Alsophis portoricensis Venom Gland Transcriptome

A total of 31,014,337 paired-end reads (100 bps) from Illumina® HiSeq 2000 sequencing were produced from the *A. portoricensis* venom gland cDNA library. Of these, 29,462,311 high-quality (phred scores greater than 30 where the probability of an incorrect base call is 1 in 1,000 and base call accuracy is 99.9%) paired-end reads were used for the gland transcriptome assemblies. The Trinity *de novo* assembly generated 123,411 contigs with an average length of

1,243 bases and an N50 of 2,279 bases. Seventy percent of overlapping paired-end reads were able to be merged, and these merged reads were used as input for the Extender assembly. Extender assembled a total of 496 contigs. After ORFpredictor was used to predict CDS from all contigs and redundant sequences from the multiple assemblies were removed with CD-HIT-EST, a total of 94,394 non-redundant contigs were assembled. Of these, 4,205 were expressed transcripts with a FPKM (fragments per kilobase of transcript per million mapped reads) value of >1, and 50 were full-length (complete CDS) venom protein transcripts that shared sequence similarity to other venom proteins in public databases and included a complete signal peptide sequence. The most abundantly expressed transcripts within the *A. portoricensis* venom gland were for venom proteins (Fig. 19). The venom gland of *A. portoricensis* demonstrated an overall higher expression of toxin transcripts than the *A. prasina* venom gland, with the top twenty expressed transcripts being toxin transcripts. Only nine of the top twenty expressed transcripts within the *A. prasina* venom gland transcriptome were toxin transcripts (Fig. 20).

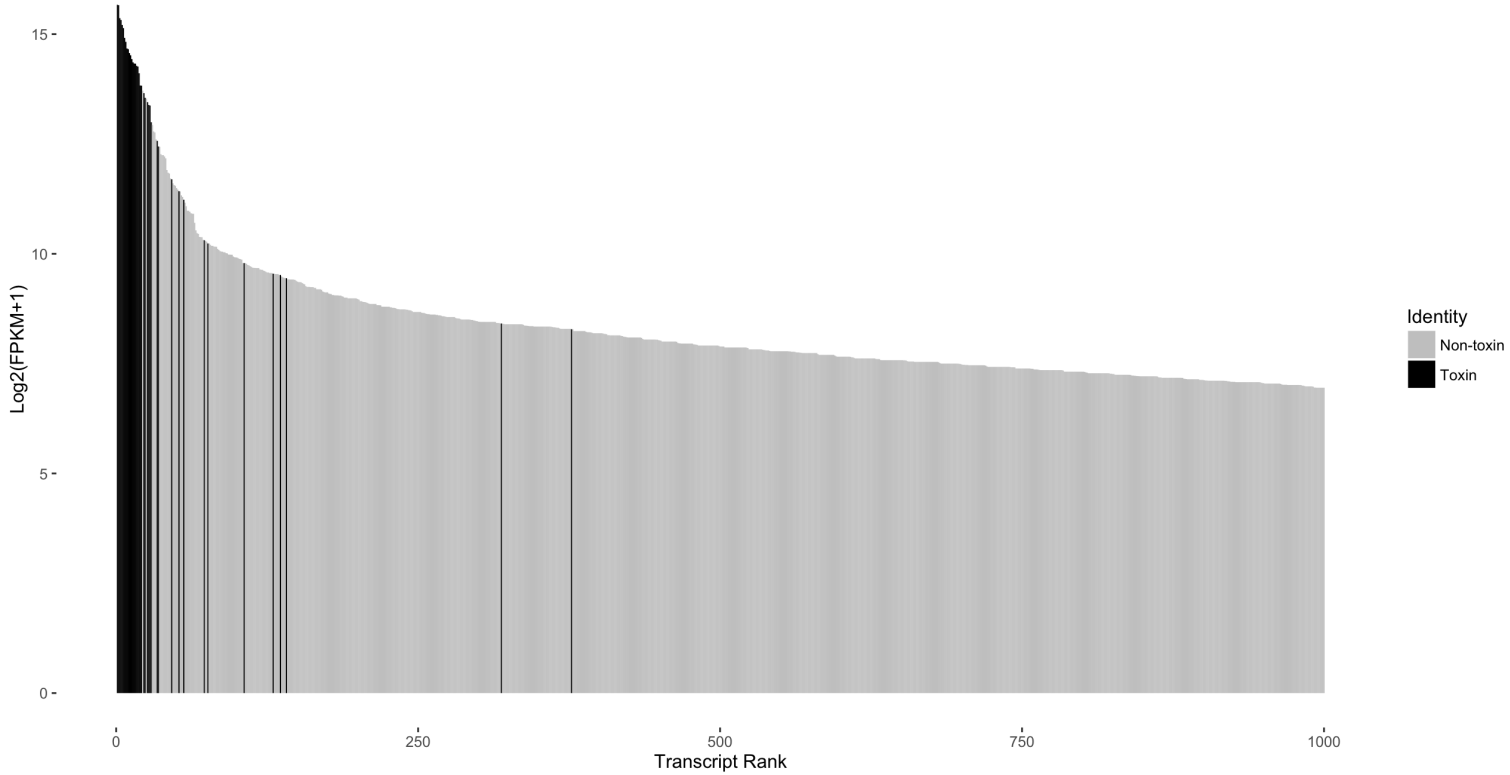


Figure 19. Venom gland tissue from *Alsophis portoricensis* demonstrates high expression of toxin transcripts. Toxins (shaded) make up the majority of the most highly expressed genes.

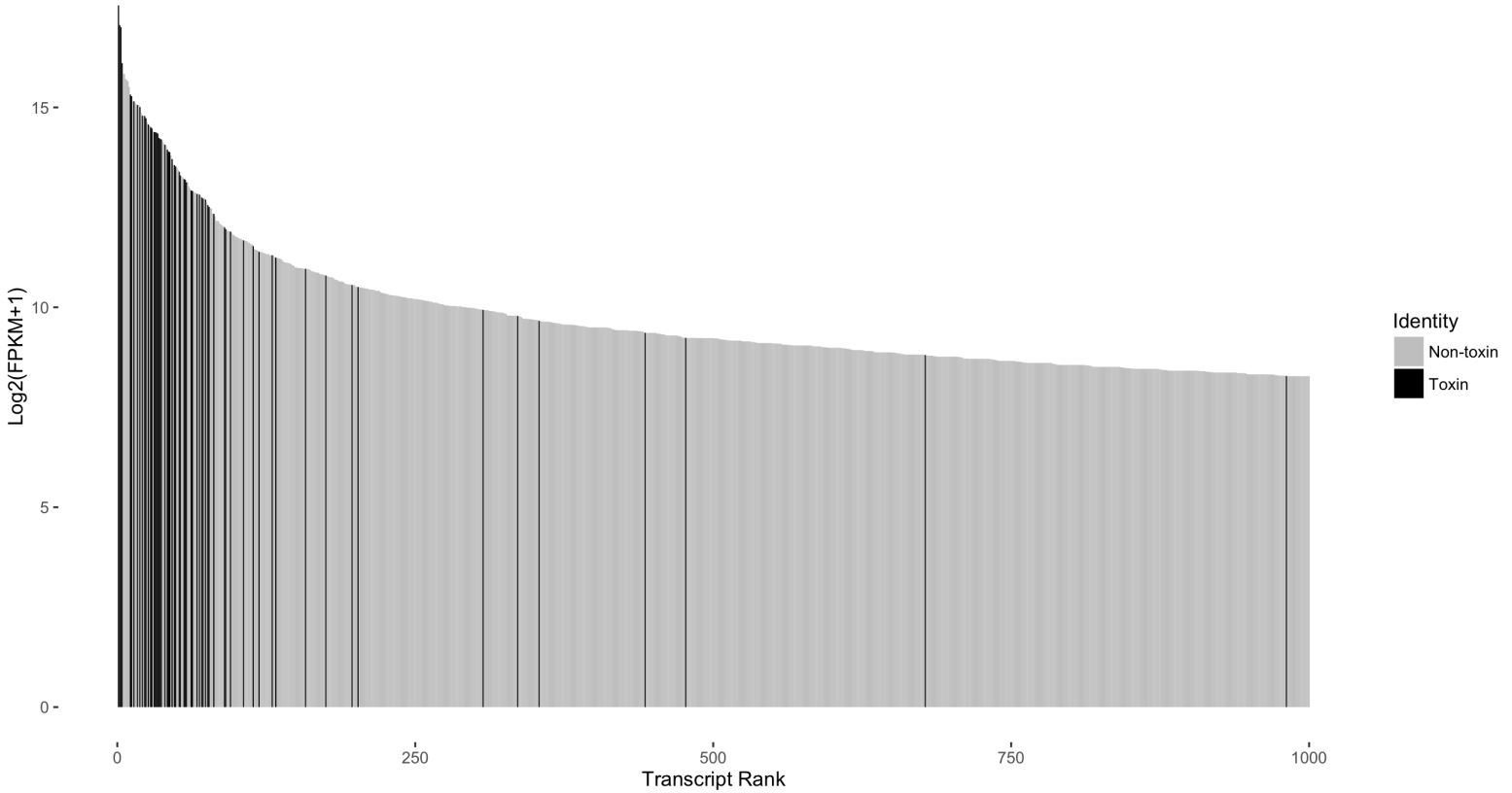


Figure 20. Venom gland tissue from *Ahaetulla prasina* demonstrates high expression of toxin transcripts. Toxins (shaded) make up the majority of the most highly expressed genes.

Eleven venom protein superfamilies were identified within the *A. portoricensis* venom gland transcriptome, and these venom protein superfamilies included SVMPs (PIII), cysteine-rich secretory proteins (CRiSPs), C-type lectins (CTLs), C-type natriuretic peptides (NPs), 3FTxs, waprins, phospholipase Bs (PLBs), venom endothelial growth factors (VEGFs), ficolins, 5'nucleases (NUCs), and a phosphodiesterase (PDEs) (Fig. 21 and Table 9). There were partial transcripts present for matrix metalloproteinases and a hyaluronidase, but because these sequences were not full length, it was therefore unknown if these transcripts encode complete functional protein products, and they were not included in the venom gland toxin transcriptome. The number of venom protein superfamilies identified within the *A. portoricensis* venom gland was similar to that been previously reported for rear-fanged snake venom gland transcriptomes (McGivern et al. 2014), demonstrating that mRNA in venom gland tissue kept at -80 °C in RNAlater for five years is quite stable and can yield a complete venom gland transcriptome assembly.

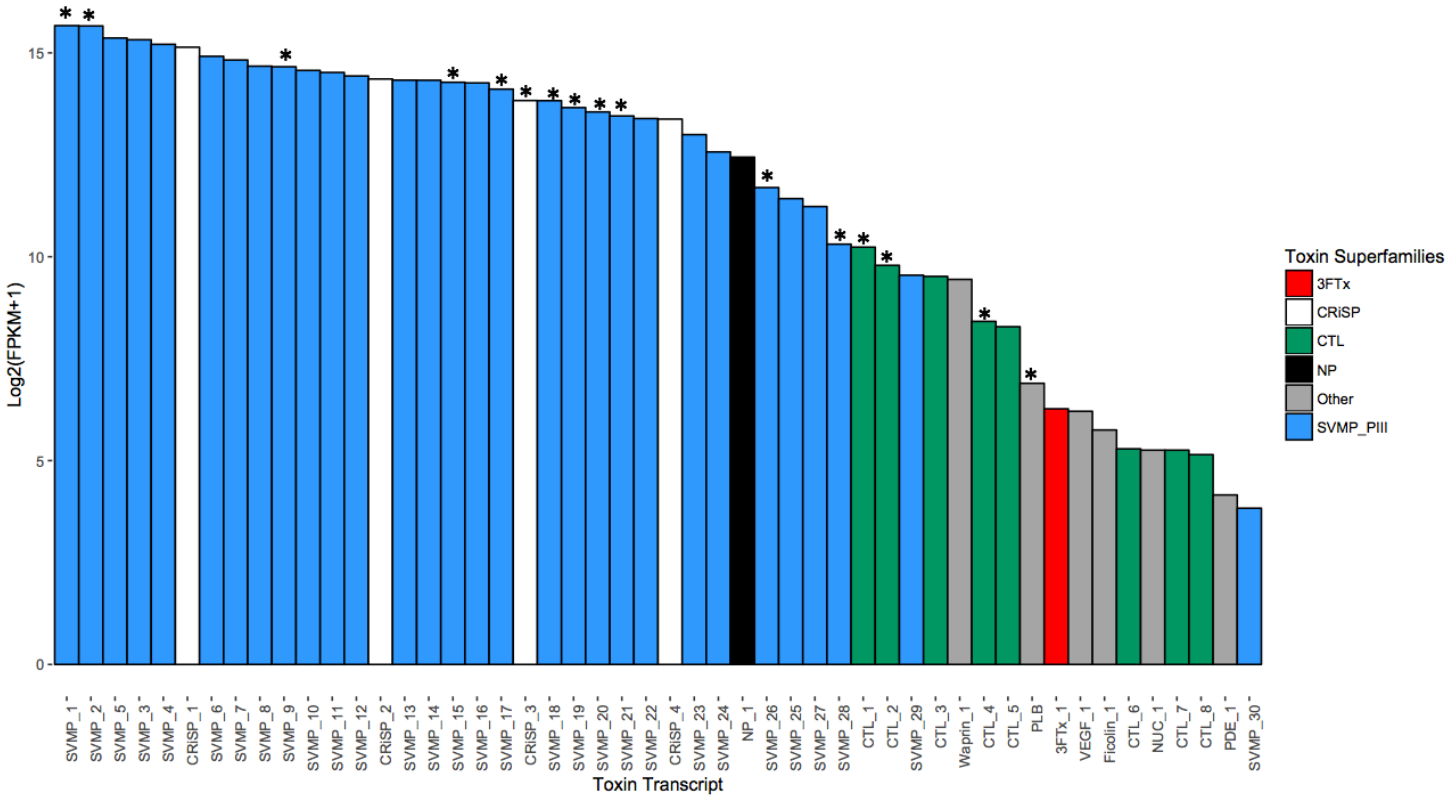


Figure 21. Venom gland tissue from *Alsophis portoricensis* demonstrates high expression of metalloproteinase transcripts. Relative expression levels of specific toxin transcripts; note that metalloproteinases dominate the transcriptome. Venom proteins that were detected within the venom and matched to a toxin transcript are marked with an asterisk. Abbreviations: 3FTx = three-finger toxin, AChE = acetylcholinesterase, CTL = C-type lectin, CRiSP = cysteine-rich secretory protein, KUN = kunitz-type protease inhibitor, PLA2_inhib = phospholipase A_2 inhibitor, PLB = phospholipase B, SVMP = snake venom metalloproteinase (P-III), VEGF = venom endothelial growth factor, and VF = venom factor.

Table 9. List of all full-length (complete CDS) venom protein transcripts from *Alsophis portoricensis* venom gland. Transcripts assembled by Extender are marked with an asterisk; all other transcripts were assembled using Trinity. Abbreviations as in Fig. 21.

Rank	Full-length Toxin Transcript ID	TPM (Transcripts Per Million)	FPKM (Fragments Per Kilobase of transcript per Million mapped reads)	Percentage of Toxin Reads
1	SVMP_1*	18513	26025	7.61%
2	SVMP_2*	18410	25881	7.57%
3	SVMP_5*	14996	21082	6.16%
4	SVMP_3	14567	20479	5.99%
5	SVMP_4	13472	18939	5.54%
6	CRiSP_1	12824	18027	5.27%
7	SVMP_6*	10989	15448	4.51%
8	SVMP_7*	10319	14507	4.24%
9	SVMP_8*	9294	13065	3.82%
10	SVMP_9	9201	12935	3.78%
11	SVMP_10*	8646	12155	3.55%
12	SVMP_11*	8351	11739	3.43%
13	SVMP_12*	7864	11055	3.23%
14	CRiSP_2*	7472	10504	3.07%
15	SVMP_13*	7330	10305	3.01%
16	SVMP_14*	7316	10285	3.00%
17	SVMP_15*	7067	9935	2.90%
18	SVMP_16*	6990	9826	2.87%
19	SVMP_17*	6288	8840	2.58%
20	CRiSP_3*	5181	7283	2.13%
21	SVMP_18*	5170	7269	2.12%
22	SVMP_19*	4594	6458	1.88%
23	SVMP_20*	4269	6001	1.75%
24	SVMP_21*	3985	5602	1.63%
25	SVMP_22*	3815	5363	1.56%
26	CRiSP_4	3778	5311	1.55%
27	SVMP_23*	2902	4080	1.19%
28	SVMP_24*	2163	3041	0.89%
29	NP_1	1979	2782	0.81%
30	SVMP_26*	1180	1658	0.48%
31	SVMP_25*	978	1374	0.40%
32	SVMP_27*	854	1201	0.35%

Table 9 Continued.

33	SVMP_28*	450	633	0.18%
34	CTL_1	428	602	0.17%
35	CTL_2*	314	442	0.12%
36	SVMP_29*	265	373	0.10%
37	CTL_3	260	366	0.10%
38	Waprin_1	247	348	0.10%
39	CTL_4	121	170	0.05%
40	CTL_5*	110	155	0.04%
41	PLB	42	59	0.01%
42	3FTx_1	27	38	0.01%
43	VEGF_1	26	36	0.01%
44	Ficolin_1	19	26	<0.01%
45	CTL_6	13	19	<0.01%
46	CTL_7	13	19	<0.01%
47	NUC_1	13	19	<0.01%
48	CTL_8	12	17	<0.01%
49	PDE_1	6	8	<0.01%
50	SVMP_30*	5	7	<0.01%

The P-III metalloproteinases were the most abundant venom protein superfamily in the *A. portoricensis* venom gland transcriptome and made up 70% of the venom transcript reads (Table 9). There were 30 SVMP PIII isoforms. The Extender assembler was much more successful at assembling these multiple SVMP isoforms and assembled 27/30 SVMP isoforms, whereas Trinity was only able to assemble 3 isoforms. Metalloproteinases belonging to the PIII class contain a disintegrin-like domain and a cysteine-rich domain, in addition to the metalloproteinase domain shared by all proteins within this superfamily (Fox and Serrano 2010; Casewell et al. 2011); only SVMPs belonging to the PIII class have been identified in rear-fanged snakes (Peichoto et al. 2011b; Weldon and Mackessy 2012; McGivern et al. 2014). These proteins are largely responsible for hemorrhagic activity secondary to degradation of basement membrane and

adhesion proteins (Fox and Serrano 2010), and are the major venom toxins within some rear-fanged snake venoms. These toxins can produce systemic symptoms upon envenomation, and snakes with these venom proteins are potentially hazardous to human health (Kamiguti et al. 2000; da Rocha and Furtado 2007; Peichoto et al. 2011b; Weldon and Mackessy 2012).

Other abundant venom protein superfamily transcripts included CTLs (11% of toxin reads) and CRiSPs (10%) (Table 9). The venom protein superfamily with the second most diversity within the *A. portoricensis* venom gland transcriptome was CTLs with 8 different isoforms, followed by CRiSPs that had 4 isoforms. Cysteine-rich secretory proteins are commonly found in many venoms, but their biological role in envenomation remains unclear (Mackessy and Heyborne 2010). C-type lectins are also unusual in that these transcripts are expressed in many venom glands, even though these proteins are not always components of some venoms (Vonk et al. 2013; McGivern et al. 2014). There was only one full-length 3FTx identified within the *A. portoricensis* venom gland transcriptome, unlike the *A. prasina* venom gland, which expressed 10 different 3FTx transcript isoforms.

Ahaetulla prasina Venom Gland Transcriptome

For the *A. prasina* venom gland cDNA library, a total of 32,316,182 paired-end reads (100 bps) were obtained from Illumina® HiSeq 2000 sequencing. Of these, 30,512,862 high-quality (phred scores greater than 30) paired-end reads were used for the venom gland transcriptome assemblies. The Trinity *de novo* assembly generated 148,863 contigs with an average contig length of 842 bases

and a N50 of 1752 bases. Merging of overlapping pair-end reads resulted in 68% of reads able to be merged, and these merged reads were used as input for the Extender assembly. Extender assembled a total of 524 contigs. After ORFpredictor was used to obtain predicted CDS from all contigs and redundant sequences from the multiple assemblies removed with CD-HIT-EST, a final total of 129,359 non-redundant contigs were assembled. Of these, 6,533 were expressed transcripts with a FPKM value of >1, and 69 were full-length venom protein transcripts that shared sequence similarity to other venom proteins and included a complete signal peptide sequence. This is 19 more toxin transcripts than what was observed for the *A. portoricensis* venom gland transcriptome. Although the most abundantly expressed transcripts were for venom proteins (Fig. 20), there were more highly expressed non-toxin transcripts than what was seen within the *A. portoricensis* venom gland transcriptome. However, sixty-two of the 69 full-length venom proteins were still within the top 1,000 expressed transcripts.

The 69 full-length venom proteins within the *A. prasina* venom gland comprised 12 venom protein superfamilies (Fig. 22 and Table 10). Transcripts were identified for 3FTxs, SVMPs (PIII), acetylcholinesterases (AChEs), CTLs, ficolins, CRiSPs, PLBs, venom factors (VFs), VEGFs, waprins, kunitz-type protease inhibitors (KUNs), and phospholipase A_2 inhibitors. The most highly expressed venom protein superfamily within the *A. prasina* venom gland was P-III metalloproteinases, 62% of the reads for toxin transcripts belonged to metalloproteinase transcripts. Interestingly, it appears that for the *A. portoricensis*

venom gland, the most highly expressed SVMP isoforms are very close in expression level, whereas for *A. prasina*, the gene expression within the venom gland is dominated primarily by several highly abundant SVMPs and then others that are expressed at slightly lower levels (Table 10). Metalloproteinases also had the highest number of present isoform transcripts with 39 identified isoforms within the *A. prasina* venom gland. Most of the full-length metalloproteinase isoforms (38/39) were assembled with Extender (Table 10).

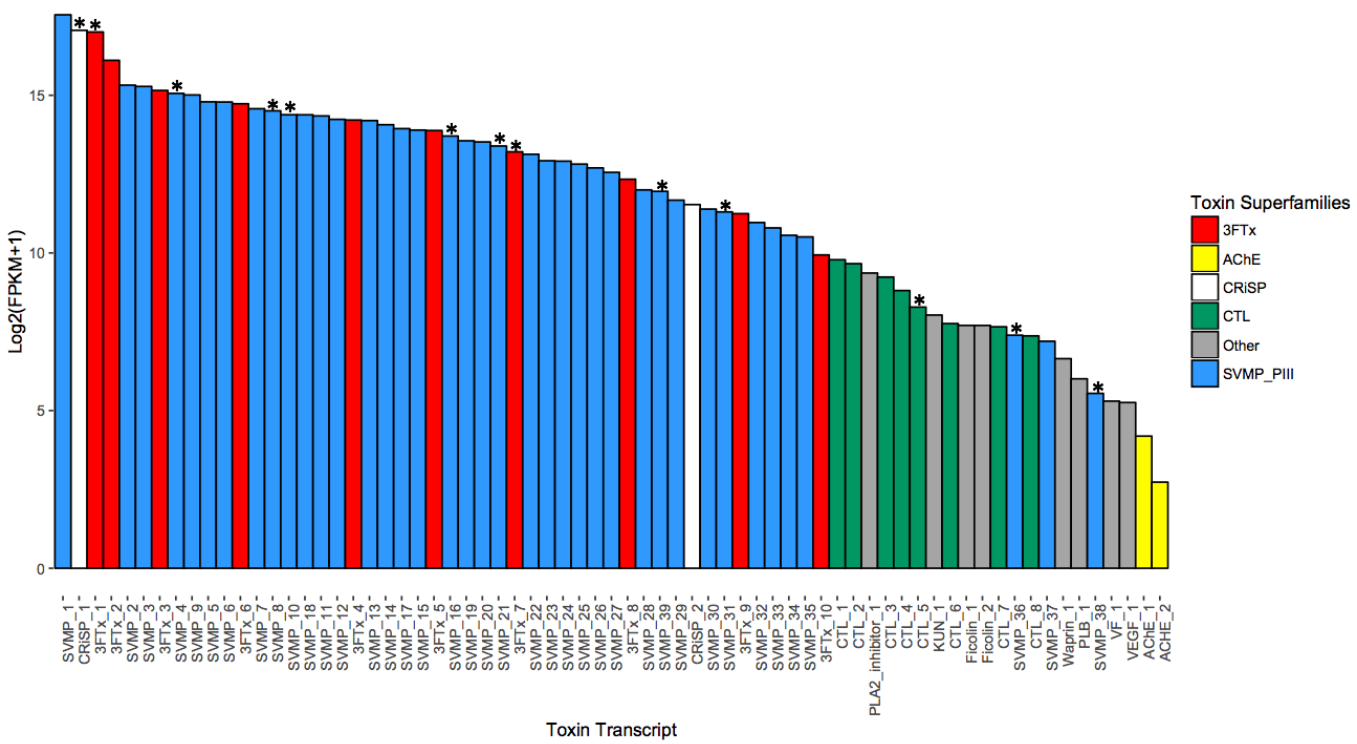


Figure 22. Venom gland tissue from *Ahaetulla prasina* demonstrates high expression of metalloproteinase transcripts. Relative expression levels of specific toxin transcripts; note that metalloproteinases dominate the transcriptome. Venom proteins that were detected within the venom and matched to a toxin transcript are marked with an asterisk. Abbreviations: 3FTx = three-finger toxin, AChE = acetylcholinesterase, CTL = C-type lectin, CRiSP = cysteine-rich secretory protein, KUN = kunitz-type protease inhibitor, PLA2_inhib = phospholipase A_2 inhibitor, PLB = phospholipase B, SVMP = snake venom metalloproteinase (P-III), VEGF = venom endothelial growth factor, and VF = venom factor.

Table 10. List of all full-length (complete CDS) venom protein transcripts from the *Ahaetulla prasina* venom gland. Transcripts assembled by Extender are marked with an asterisk, all other transcripts were assembled using Trinity. Abbreviations as in Fig. 22.

Rank	Full-length Toxin Transcript ID	TPM (Transcripts Per Million)	FPKM (Fragments Per Kilobase of transcript per Million mapped reads)	Percentage of Toxin Reads
1	SVMP_1	41525	96148	16.39%
2	CRiSP_1	29502	68308	11.64%
3	3FTx_1*	28451	65876	11.23%
4	3FTx_2*	15254	35320	6.02%
5	SVMP_2*	8842	20473	3.49%
6	SVMP_3*	8595	19901	3.39%
7	3FTx_3	7871	18225	3.10%
8	SVMP_4*	7378	17083	2.91%
9	SVMP_9*	7128	16505	2.81%
10	SVMP_5*	6128	14189	2.41%
11	SVMP_6*	6103	14131	2.40%
12	3FTx_6*	5874	13602	2.31%
13	SVMP_7*	5267	12196	2.07%
14	SVMP_8*	5030	11647	1.98%
15	SVMP_10*	4613	10683	1.82%
16	SVMP_18*	4607	10668	1.81%
17	SVMP_11*	4495	10409	1.77%
18	SVMP_12*	4167	9650	1.64%
19	3FTx_4	4106	9507	1.62%
20	SVMP_13*	4058	9396	1.60%
21	SVMP_14*	3705	8579	1.46%
22	SVMP_17*	3401	7876	1.34%
23	SVMP_15*	3293	7625	1.30%
24	3FTx_5	3261	7552	1.28%
25	SVMP_16*	2889	6690	1.14%
26	SVMP_19*	2602	6026	1.02%

Table 10 continued.

27	SVMP_20*	2538	5877	1.00%
28	SVMP_21*	2319	5371	0.91%
29	3FTx_7	2043	4731	0.80%
30	SVMP_22*	1935	4481	0.76%
31	SVMP_23*	1677	3884	0.66%
32	SVMP_24*	1663	3852	0.65%
33	SVMP_25*	1558	3609	0.61%
34	SVMP_26*	1433	3320	0.56%
35	SVMP_27*	1300	3010	0.51%
36	3FTx_8	1117	2587	0.44%
37	SVMP_28*	884	2047	0.34%
38	SVMP_39*	856	1983	0.33%
39	SVMP_29*	700	1633	0.27%
40	CRiSP_2*	640	1481	0.25%
41	SVMP_30*	579	1341	0.22%
42	SVMP_31*	545	1262	0.21%
43	3FTx_9*	524	1214	0.20%
44	SVMP_32*	431	998	0.17%
45	SVMP_33*	383	887	0.15%
46	SVMP_34*	325	754	0.12%
47	SVMP_35*	314	728	0.12%
48	3FTx_10	211	489	0.08%
49	CTL_1	190	441	0.07%
50	CTL_2	174	404	0.06%
51	PLA2_inhibitor_1	142	329	0.05%
52	CTL_3	130	301	0.05%
53	CTL_4	96	223	0.03%
54	CTL_5	67	155	0.02%
55	KUN_1	56	130	0.02%
56	CTL_6	46	108	0.01%
57	Ficolin_1	44	104	0.01%
58	Ficolin_2	44	104	0.01%
59	CTL_7	43	100	0.01%
60	SVMP_36*	36	83	0.01%
61	CTL_8	35	82	0.01%
62	SVMP_37*	31	73	0.01%
63	Waprin_1	21	50	<0.01%
64	PLB_1	13	32	<0.01%
65	SVMP_38*	10	23	<0.01%
66	VF_1	8	9	<0.01%

Table 10 continued.

67	VEGF_1	8	19	<0.01%
68	AChE_1	3	9	<0.01%
69	AChE_2	1	3	<0.01%

Other highly expressed venom protein transcripts included 3FTxs (17% of toxin reads), CTLs (9%), and CRiSPs (4%) (Fig. 22 and Table 10). There were 10 3FTx isoforms present, which was identified as the second most diversified venom protein superfamily. C-type lectins were the third most diversified venom protein superfamily with 8 isoforms.

Dominant Venom Protein Superfamilies Within Rear-Fanged Snake Transcriptomes

Transcripts for SVMPs and 3FTxs within the *A. portoricensis* and *A. prasina* toxin transcriptome were examined in detail given the abundance and diversity of these transcripts within rear-fanged snakes. It has been shown that differences in expression of these two venom protein superfamilies result in two different venom phenotypes (McGivern et al. 2014). Venom glands with high expression levels of SVMPs will result in venoms with higher proteolytic and enzymatic activity, as observed within this current study, whereas venom gland transcriptomes with high expressions of 3FTxs result in venoms that display higher toxicities and are usually neurotoxic (Mackessy et al. 2006; Rokyta et al. 2012a; Margres et al. 2013; McGivern et al. 2014) (Fig. 23).

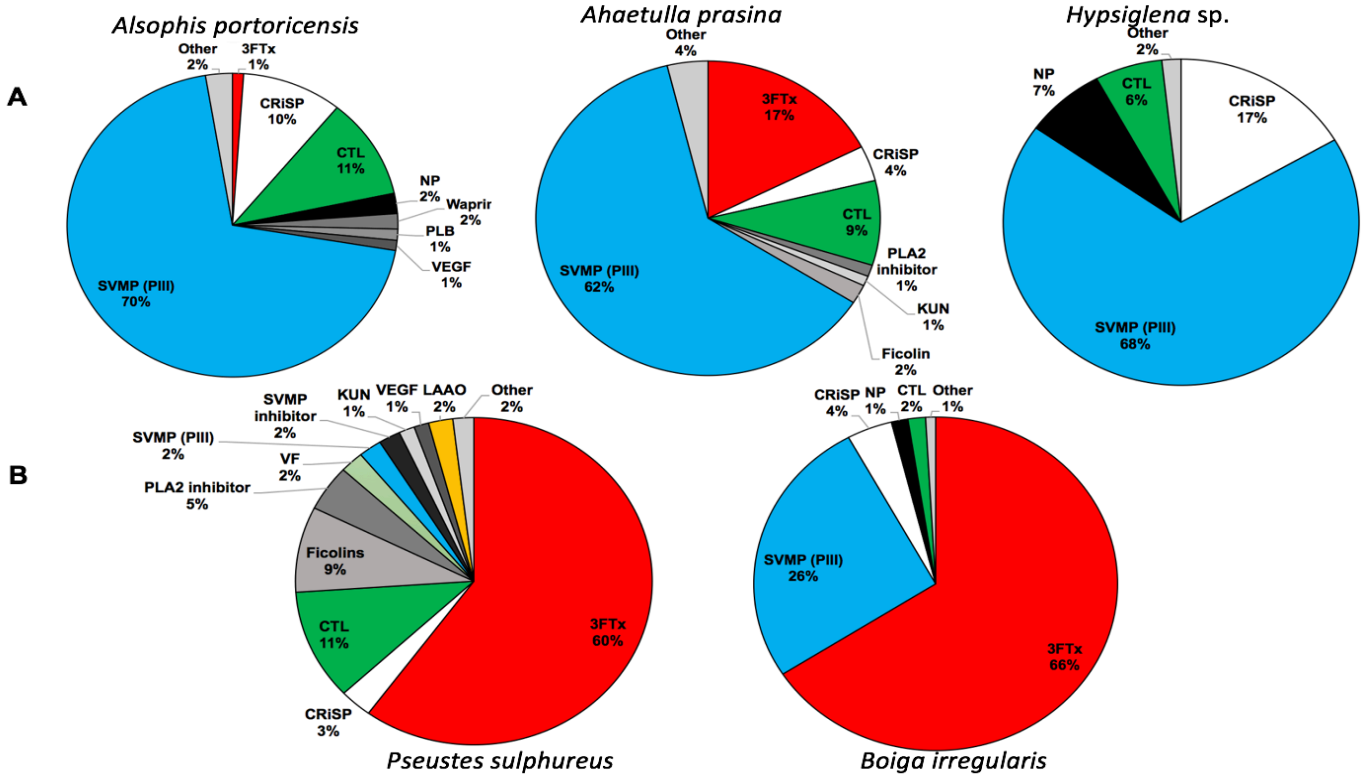


Figure 23. Rear-fanged snake venom gland transcriptomes demonstrate similarities to front-fanged viperid and elapid snakes. (A) The venom gland transcriptomes of *A. portoricensis* and *A. prasina* are viperid-like, with high abundance of metalloproteinase transcripts. (B) The elapid-like transcriptomes of *Pseustes* and *Boiga* show high abundance of three-finger toxins. Venom protein superfamilies with abundances less than 1% were grouped as "other". Abbreviations: 3FTx = three-finger toxin, CTL = C-type lectin, CRiSP = cysteine-rich secretory protein, KUN = kunitz-type protease inhibitor, LAAO = L-amino acid oxidase, NP= natriuretic peptide, PLA2 inhibitor = phospholipase A_2 inhibitor, PLB = phospholipase B, SVMP (PIII) = snake venom metalloproteinase type P-III, SVMP inhibitor = snake venom metalloproteinase inhibitor, VEGF = venom endothelial growth factor, and VF = venom factor.

For the *A. portoricensis* venom gland transcriptome, several SVMP isoforms showed a 100% identity to the short amino acid sequence obtained from N-terminal sequencing of Alsophinase, the characterized P-III SVMP from the *A. portoricensis* venom (Weldon and Mackessy 2012). Most likely the *A. portoricensis* SVMP_1 isoform corresponds to the protein Alsophinase since this is the most abundant SVMP transcript and Alsophinase was one of the most abundant SVMP within the crude venom (Weldon and Mackessy 2012). Interestingly, it appears that in the case of at least three SVMP transcripts, the stop codon for protein translation termination experienced a nucleotide substitution that allowed translation to continue into the 3'UTR (untranslated region) of the transcript. This extension produced SVMPs with an extended nine amino acid C-terminus (Fig. 24, only two isoform examples shown). Overall, both *A. portoricensis* and *A. prasina* SVMPs shared high identities. All of the *A. prasina* SVMP isoforms were over 70% identical in amino acid sequence, and all of the *A. portoricensis* SVMP isoforms were over 75% identical (Fig. 24). Alignments between the SVMPs from both venom gland transcripts showed over 69% of the SVMP amino acid sequence conserved.

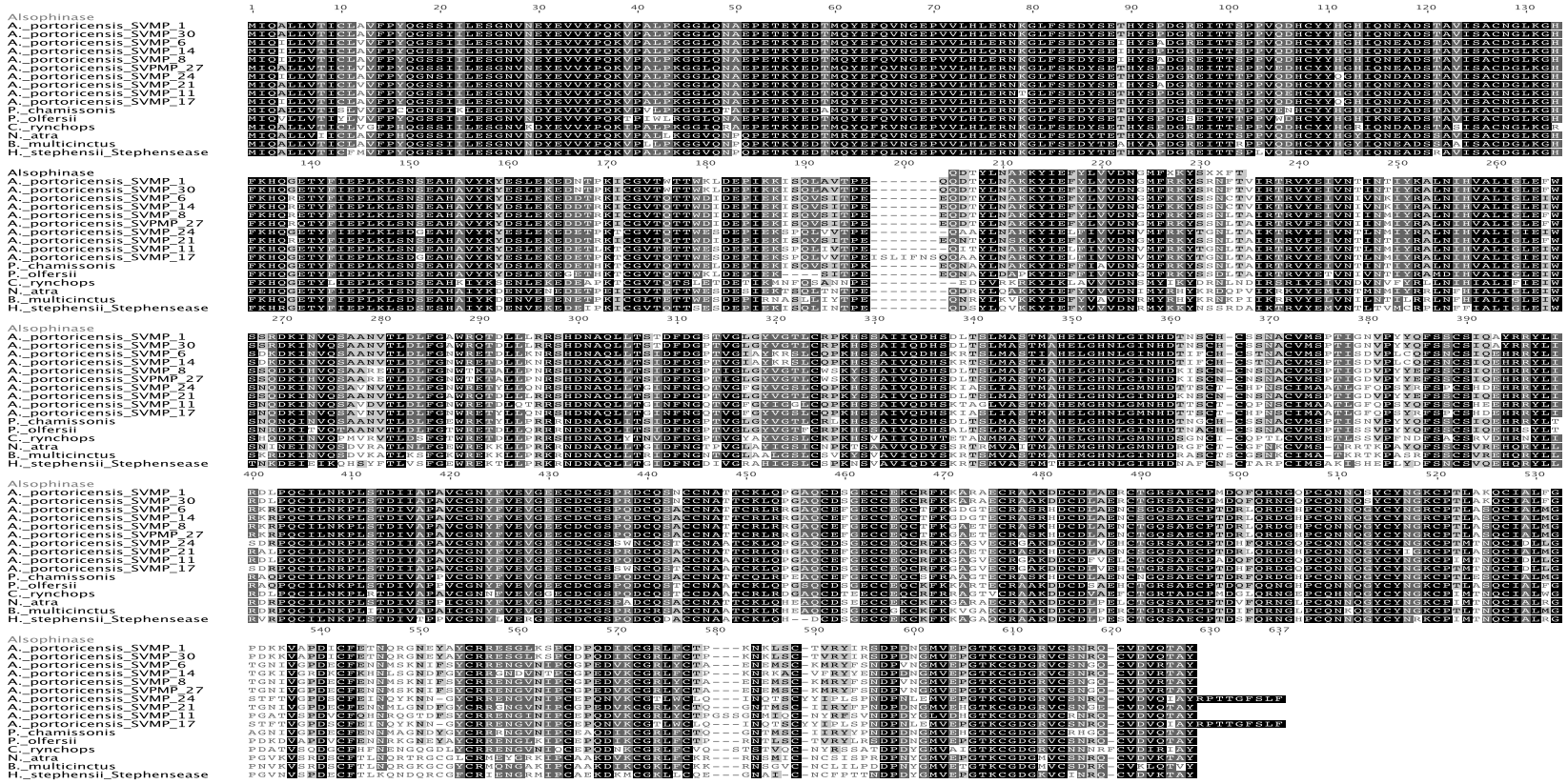


Figure 24. Snake venom metalloproteinase (PIII) sequence alignments exhibit overall similarity within the *Alsophis portoricensis* venom gland and with other rear-fanged and elapid venomous snakes. Amino acids are shaded based on shared sequence within each region. Darker shading indicates conservation of amino acid sequence. The partial amino acid sequence of the isolated and characterized metalloproteinase within *A. portoricensis* venom, Alsophinase, is also shown. Other metalloproteinase sequences were obtained from *Philodryas chamissonis* (AJB84503.1), *P. olfersii* (ACS74988.1), *Cerberus rynchops* (VM3_CERRY), *Naja atra* (VM3H_NAJAT), *Bungarus multicinctus* (VM3_BUNMU), and *Hoplocephalus stephensii* (ABQ01135.1)

Transcripts for 3FTxs did not show as much amino acid sequence conservation as what is observed for SVMPs. Within the *A. prasina* venom gland transcriptome, 3FTx isoforms shared 34.8-98.95% identity. The venom gland transcriptome of *A. portoricensis* only exhibited one 3FTx transcript. Interestingly, the one 3FTx transcript present shared low similarity with 3FTx transcripts identified within the *A. prasina* venom gland transcriptome. The single 3FTx transcript from the *A. portoricensis* venom gland transcriptome was more similar to 3FTxs from Asian Catsnakes (*Boiga* sp.), which exhibit an extended N-terminal (Fig. 25). The 3FTx sequences from the *A. prasina* venom gland were more similar (49% identical) to 3FTxs identified within venom gland transcriptomes of the *Dispholidus typus* (Boomslang) and *Thrasops jacksonii* (Jackson's Black Tree Snake), and species of cobras (*Ophiophagus hannah* and *Naja sputatrix*). This provides interesting insight into the expression and evolution of 3FTx transcripts.

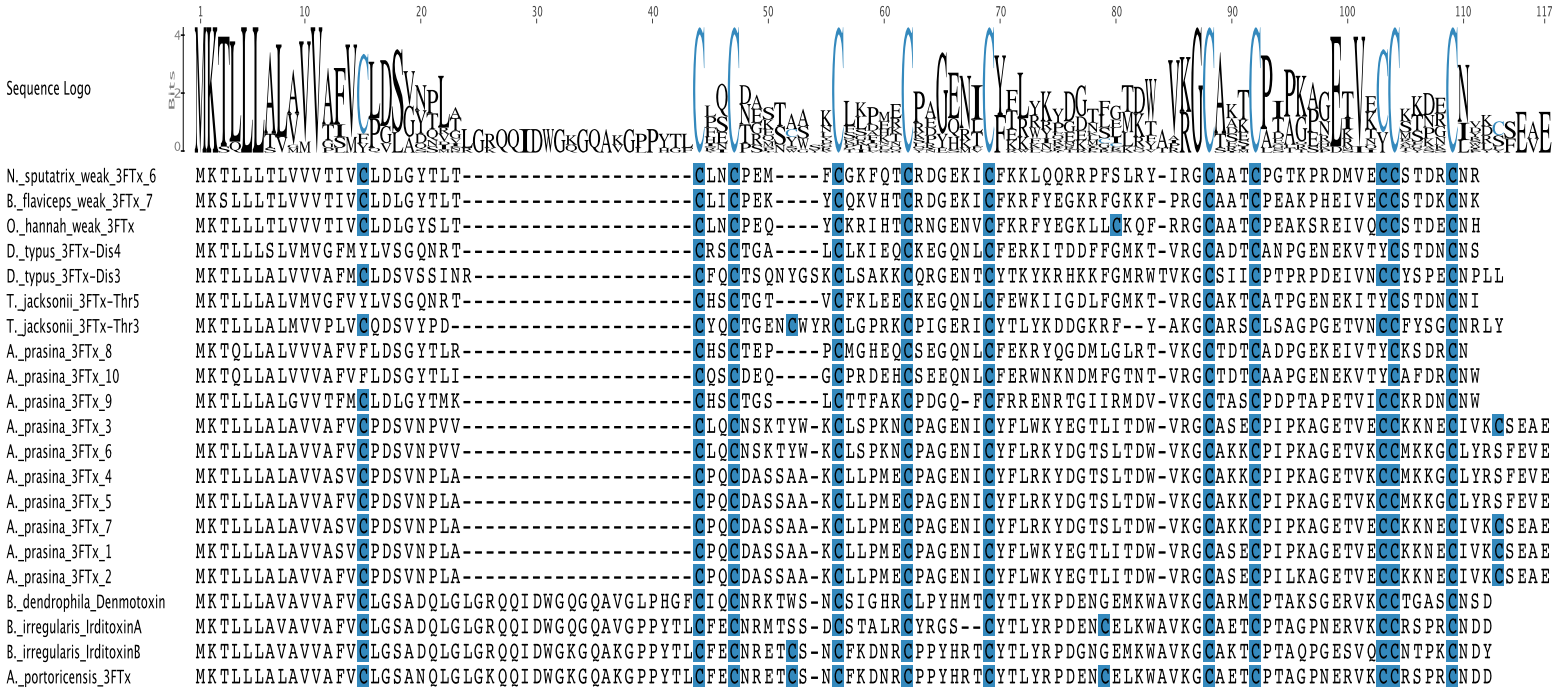


Figure 25. Sequence alignment of three-finger toxin isoforms from rear-fanged snakes and elapids. Cysteines are highlighted due to the conserved three-finger toxin structure obtained from disulphide linkages between cysteines. The five disulphide linkages characteristic of non-conventional 3FTxs are shown with solid lines. Genbank accessions include: denmotoxin (Q06ZW0.1), irditoxin subunit A (A0S864.1), irditoxin subunit B (A0S865.1), *Thrasops jacksonii* (ABU68485.1 and 3SX5_THRJA), *Dispholidus typus* (ABU68483.1 and 3SX4_DISTY), *Ophiophagus hannah* (3NO24_OPHHA), *Naja sputatrix* (3NO26_NAJSP), and *Bungarus flaviceps* (ADF50022.1).

Within the next chapter, two 3FTx transcripts were isolated from the crude venom of *A. prasina.* One of these transcripts was revealed to be 100% identical to a 3FTx transcript within this *A. prasina* venom gland transcriptome. The other transcript was 96.5% identical and the few amino acid differences appearing to be an additional isoform, possibly due to locality differences between the *A. prasina* used in each study. It is impossible to know if they were collected from the same locality due to them both being pet trade imports. This demonstrates that it was possible to acquire the same toxin transcripts from crude venom that are present within the venom gland transcriptome.

Alsophis portoricensis
Venom Proteome

Overall toxin gene expression levels for the most abundant venom protein superfamilies closely resembled that observed for the venom proteome of *A. portoricensis* (Fig. 26A-B), although not all isoforms for each superfamily were found in the venom proteome. Only 16 of the 50 venom protein transcripts were identified in *A. portoricensis* crude venom LC-MS/MS (Fig. 21). Similar ratios of transcripts and translated protein products were reported for the venom gland transcriptomes and proteomes of the rear-fanged snakes *Hypsiglena* sp. and *Boiga irregularis* (McGivern et al. 2014).

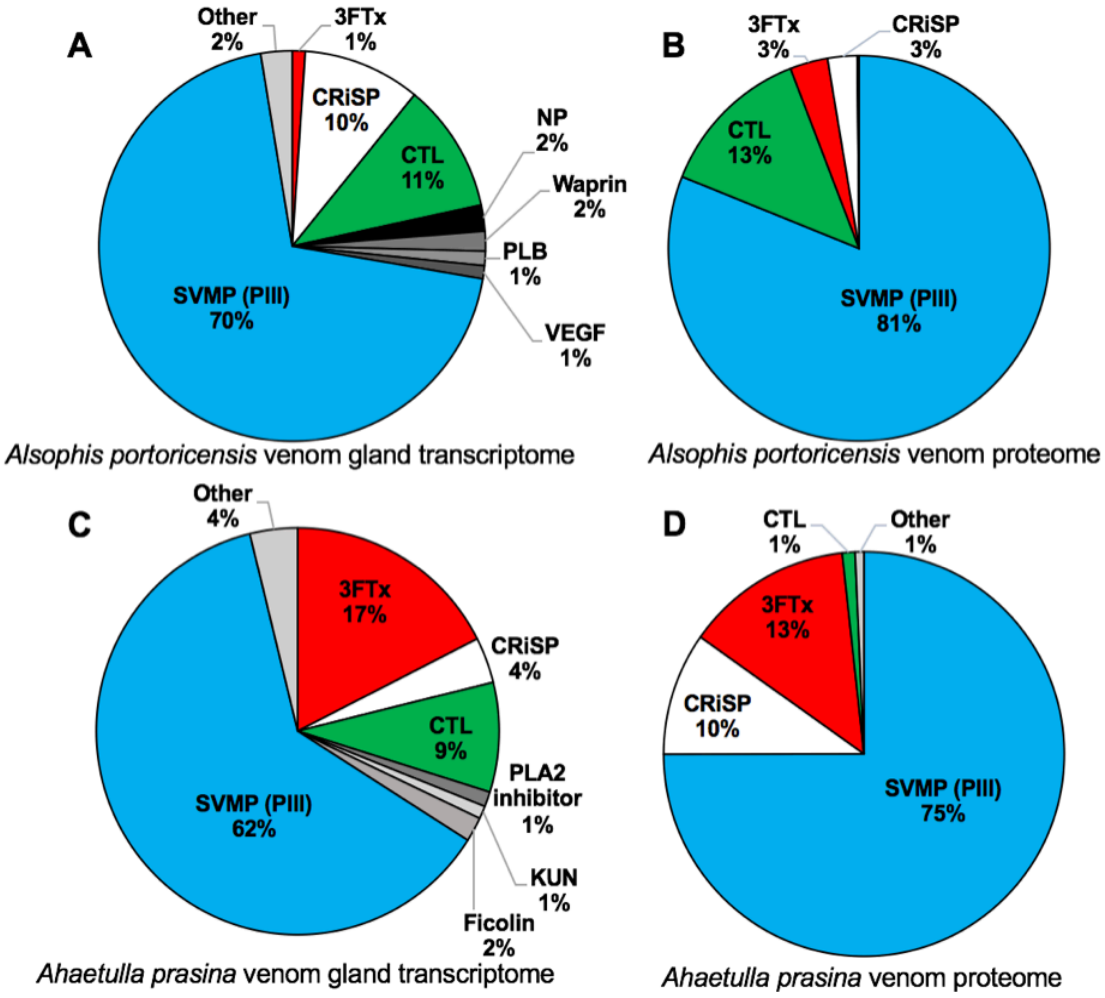


Figure 26. Abundances of venom protein superfamilies within two rear-fanged snake venom gland transcriptomes and venom proteomes. Transcripts belonging to each venom protein superfamily within the *Alsophis portoricensis* venom gland were quantified (A), in addition to venom proteins within this venom (B). Transcripts belonging to each venom protein superfamily within the *Ahaetulla prasina* venom gland were also quantified (C), in addition to venom proteins within this venom (D). Venom protein superfamilies with abundances less than 1% were grouped within the category “other”. Abbreviations as in Fig. 23.

Using a custom rear-fanged snake toxin transcriptome (see Materials and Methods) as a reference to identify LC-MS/MS peptide fragments from trypsin digested venom, six venom protein superfamilies were identified within *A. portoricensis* venom. These venom protein superfamilies include SVMP PIIIs, CTLs, 3FTxs, CRiSPs, venom factors, and PLBs. Using the available databases (Mascot, Sequest, and X! Tandem), eight venom protein superfamilies were identified. These included SVMPs PIIIs, CTLs, 3FTxs, CRiSPs, matrix metalloproteinases, PLA_2 inhibitors, venom factors, and C-type natriuretic peptides. The use of the larger database allowed for more venom protein superfamilies to be identified because the toxin transcriptome reference generated from the *A. portoricensis* venom gland (among other rear-fanged venomous snake venom gland transcriptomes) only included full-length (complete CDS) transcripts. It is possible that due to mRNA degradation or issues with next-generation sequencing assemblies that some toxin transcripts within the *A. portoricensis* venom gland were not assembled (Macrander et al. 2015). However, more overall peptide fragments were identified using the custom rear-fanged snake venom gland transcriptome database. Without the *A. portoricensis* toxin transcriptome, only 199 peptide spectra were identified, whereas 1,138 peptide spectra were identified with the *A. portoricensis* rear-fanged snake toxin transcriptome reference. This allows for better quantification of venom protein superfamilies within the venom because fewer peptide spectra are excluded (Table 11). Venom protein superfamilies within the *A. portoricensis* venom proteome were quantified using the NSAF approach with the used *A.*

portoricensis venom gland toxin transcriptome reference for LC-MS/MS peptide spectra numbers. Metalloproteinases (PIII) were found to be the most abundant (81%), followed by CTLs (13%), CRiSPs (3%) and 3FTxs (3%) (Fig. 26).

Table 11. All MS/MS identified venom proteins from *Alsophis portoricensis* venom, using the custom rear-fanged snake venom gland toxin combined transcriptomes as a reference. All peptide fragments for venom proteins that had at least 10 total spectra are shown. Abbreviations: CRiSP = cysteine-rich secretory protein, CTL = C-type lectin, and SVMP (PIII) = snake venom metalloproteinase type PIII.

Protein Identity (transcript)	MS/MS-derived Peptide Sequences	Spectra	Protein superfamily
A.port.SVMP_21	ALNIHVALIGLEFWSSQDK ASKHDcDLAENcSGQSAEcPTDRLQR cGDGRVcSNGEcVDVQTAY DGHPcQNNQGYcYIGR KYIEFYLVVDNGMFRK LYcTQGNTMScIIR RGNGVNIPcGPEDVKcGR RSHDNAQLLTSIDFDGPTVGLGYVGTL cRPK TRVFEIVNTINTIYR YFPNDPDNGMVEHGTK YIEFYLVVDNGMFR YSSAIVQDHSDLTSLMASTMAHELGHN LGINHDK	213	SVMP PIII
A.port.SVMP_19	ALNIHVALIGLEFWSSQDKIHVQSAAK ASKHDcDLAENcSGQSAEcPTDRLQR cGDGRVcSNGEcVDVQTAY DGHPcQNNQGYcYIGR DRPQCILNKPLSTDIVAPAVcGNYFVEV GEEcDcGSPR GNGVNIPcGPEDVKcGR HDcDLAENcSGQSAEcPTDR KYIEFYLVVDNGMFR LYcTQGNTMScIIR RGNGVNIPcGPEDVKcGR RSHDNAQLLTSIDFDGPTIGLGYVGTLc WLK TRVFEIVNIINMIYR VFEIVNIINMIYRALNIHVALIGLEFWSSQ DKIHVQSAAK YFPNDPDNGMVEHGTK YIEFYLVVDNGMFR	155	SVMP PIII

Table 11 continued.

A.port.SVMP_20	ALNIHVALIGLEFWSSQDKIHVQSAAR ASKHDcDLAENcSGQSAEcPTDRLQR DGHPcQNNQGYcYIGR HDcDLAENcSGQSAEcPTDR KYIEFYLVVDNGMFR LYcTQGNTMScIIR RGNGVNIPcGPEDVKcGR TRVFEIVNIINMIYR VFEIVNIINMIYRALNIHVALIGLEFWSSQ DKIHVQSAAR YFPNDPDNGMVEHGTK YIEFYLVVDNGMFR YSSAIVQDHSDLTSLMASTMAHELGHN LGINHDK	115	SVMP PIII
A.port.SVMP_2	ALNIHVALIGLEFWSSR cPIMTNQcIDLLGPGATVSPDVcFQHNR DDcDFVELcTGQSAEcPADQFQR FSVNDPDYGLVDHGTKcGDGR GAKDDcDFVELcTGQSAEcPADQFQR KYIEFYLVVDNGmFRK LYcTPGSSGNMIQcNYR RENGINIPcEPQDVK RSHDNAQLLTSTDFDGPTVGLGYVGTL cRPK TRVYEIVNTINTIYK TRVYEIVNTINTIYKALNIHVALIGLEFWS SR VYEIVNTINTIYK YIEFYLVVDNGMFR	101	SVMP PIII
A.port.SVMP_18	ALNIHVALIGLEFWSSQDKIHVQSAAR ASKHDcDLAENcTGQSAEcPTDR cPTLASQcIALMGTGK DGHPcQNNQGYcYNGR HDcDLAENcTGQSAEcPTDR KYIEFYLVVDNGMFRK TRVFEIVNIINMIYR VFEIVNIINMIYRALNIHVALIGLEFWSSQ DKIHVQSAAR YIEFYLVVDNGMFR YSSAIVQDHSDLTSLMASTMAHELGHN LGINHDK YYENDPDNGMVEPGTK	97	SVMP PIII

Table 11 continued.

A.port.SVMP_1	ALNIHVALIGLEFWSSRDK KVAPDIcFETNQR KYIEFYLVVDNGMFRK LQPGAQcDSGEccEK NGQPcQNNQSYcYNGK RSHDNAQLLTSTDFDGSTVGLGYVGTL cRPK TRVYEIVNTINTIYKALNIHVALIGLEFWS SR VAPDIcFETNQR VYEIVNTINTIYK YIEFYLVVDNGMFR	90	SVMP PIII
A.port.SVMP_28	ALNIHVALIGLEFWSSRDK KVAPDIcFETNQR KYIEFYLVVDNGMFR LQPGAQcDSGEccEK NGQPcQNNQSYcYNGK RSHDNAQLLTSTDFDGPTVGLGYVGTL cRPK TRVYEIVNTINTIYKALNIHVALIGLEFWS SR VAPDIcFETNQR VYEIVNTINTIYK YIEFYLVVDNGMFR	81	SVMP PIII
A.port.SVMP_26	cPIMTNQcIDLLGPGATVSPDVcFQHNR DDcDFVELcTGQSAEcPADQFQR DGQPcQNNQGYcYNGK FSVNDPDYGLVDHGTKcGDGR GAKDDcDFVELcTGQSAEcPADQFQR INVQSAVDVTLDLFGNWR KYIELFIVVDNVMFR LYcTPGSSGNMIQcNYR RENGINIPcEPQDVK RSHDNAQLLTGINFNGQTVGFGYIGGLc QPK TRVYEIVNTLNMIYR VYEIVNTLNMIYR YIELFIVVDNVMFR	78	SVMP PIII
A.port.SVMP_9	ALNIHVALIGLEFWSSQDKIHVQSAAK ASKHDcDLAENcSGQSAEcPTDRLQR cGDGRVcSNGQcVDVR DGHPcQNNQGYcYNGR HDcDLAENcSGQSAEcPTDR KYIEFYLVVDNGMFR RENGVNIPcGPEDVK RSHDNAQLLTSIDFDGPTIGLGYVGTLc WLK TRVFEIVNIINMIYR YIEFYLVVDNGMFR YSSAIVQDHSKR	65	SVMP PIII

Table 11 continued.

A.port.SVMP_17	DDcDLVEHcTGR DGQPcQNNQGYcYNGK GAKDDcDLVEHcTGR KYIELFIVVDNVMFRK RENGVNIPcEPQNVK SAEcPTDHFQRDGQPcQNNQGYcYNGK SHDNAQLLTGINFNGQTVGFGYVGSLcQPK TRVYEIVNTLNMIYR VYEIVNTLNMIYR YIELFIVVDNVMFR	40	SVMP PIII
A.port.SVMP_15	ASKHDcDLAENcTGQSAEcPTDR cGDGRVcSNGQcVDVR DGHPcQNNQGYcYNGR HDcDLAENcTGQSAEcPTDR KYIEFYLVVDNGMFKK RENGVNIPcGPEDVK SHDNAQLLTSIDFDGPTVGIAYKR TRVYEIVNIVNK VYEIVNIVNKIYR	23	SVMP PIII
A.port.CTL_1	FcMEQGKGGHLASLGSIEEGNFVGK GGHLASLGSIEEGNFVGKLTFKR KNWDDAERFcMEQGK NWDDAERFcMEQGKGGHLASLGSIEEGNFVGK WAEYQKWHNYNcDFTIPFIcK WHNYNcDFTIPFIcK WSDGSTIVYENWHPLQSRK	17	CTL
A.port.CTL_2	FcMEQGKGGHLASLGSIEEGNFVGK GGHLASLGSIEEGNFVGKLTFKR KNWDDAERFcMEQGK NWDDAERFcMEQGKGGHLASLGSIEEGNFVGK WAEYQKWHNYNcDFTVPFIcK	17	CTL
A.port.CTL_4	cSWQWSDGSSLSYEAWVEGSDcVIMH LKPGSIEWYSIEcK KNWDEAEKFcK SLKHPSMWIGLSNIWNK TGGHLASILSSEEGSYVANLAFK	15	CTL
A.port.CRiSP_3	MEWYSEAASNAER NEYTNcNELVQQNScQDDWTK SVNPTASNMLKMEWYSEAASNAERWA FQcAYDHSLNSER WAFQcAYDHSLNSER	14	CRiSP

Ahaetulla prasina
Venom Proteome

The *A. prasina* venom proteome was also predominately SVMPs (PIII) (75%), with 3FTxs (13%) and CRiSPs (10%) seen in moderate abundance (Fig. 26D). C-type lections were only 1% of the venom composition in comparison to the 13% observed within *A. portoricensis* venom. It appears that within *A. prasina* venom, CTLs do not play a prominent role in envenomation. Only 12 of the 69 transcripts were identified within the *A. prasina* venom proteome, a lower number compared to what was observed for *A. portoricensis* even though considerably more venom protein transcripts are present within the *A. prasina* venom gland. This is suggestive of translation regulation of multiple isoforms within a single venom protein superfamily. It is also possible that given high sequence conservation between the multiple isoforms, it is difficult for the program Scaffold to assign LC-MS/MS peptide spectra with high confidence (>99.9%) to any single isoform and/or that peptide sequences are being incorrectly assigned (Table 12). This is the challenge associated with using crude venom for trypsin digestions and LC-MS/MS (a shotgun proteomics approach) instead of purified proteins.

The use of current software platforms (Mascot, Sequest, and X! Tandem) to identify the LC-MS/MS peptide fragments from *A. prasina* venom resulted in identification of six venom protein superfamilies. However, only 70 peptide spectra were matched to proteins within these databases. Unlike for *A. portoricensis* venom LC-MS/MS, few proteins searched by these platforms were identified, as databases lacked proteins with sequence similarity to venom proteins from *A. prasina.* Six venom protein superfamilies were also identified

when interrogating a custom transcriptome database of rear-fanged snake venom gland transcriptomes, including the *A. prasina* toxin transcriptome, and 906 peptide spectra were matched to venom proteins within this custom database.

Table 12. List of all MS/MS identified venom proteins from *Ahaetulla prasina* venom using the rear-fanged snake venom gland toxin transcriptome as a reference. All peptide fragments for venom proteins that had at least 10 total spectra are shown.

Protein Identity (transcript)	MS/MS-derived Peptide Sequences	Spectra	Protein superfamily
A.pra.SVMP_10	AAPDIcFNENQKGNSYAYcR ASKHDcDMPELcTGQSAEcPIDLFQR cGDGKVcSNGEcVDVETAY EcVMSETLGFEPPYK GNSYAYcRQENGVNIPcQPK HDcDMPELcTGQSAEcPIDLFQR IKVQSASDVTLALFGIWR KAAPDIcFNENQK KRNDNAQLLTGIDFDGPVVGLGYLGTLcMPK NDNAQLLTGIDFDGPVVGLGYLGTLcMPK NKGLFAGDYSETHYSPDGR QENGVNIPcQPK VQSASDVTLALFGIWRETDLLPR YIEFYVVVDNLMYKK YSGNLSTIR	150	SVMP PIII
A.pra.SVMP_8	AAPDIcFNENLKGNSYGYcRQENGVNIPcQPQDVK ASKHDcDMPELcTGQSAEcPIDLFQR cGDGKVcSNGEcVDVQTAY EcVMSETLGFEPPYK GNSYGYcRQENGVNIPcQPQDVK HDcDMPELcTGQSAEcPIDLFQR KAAPDIcFNENLKGNSYGYcR KRNDNAQLLTGIDFDGPVVGLGYLGTLcMPK NDNAQLLTGIDFDGPVVGLGYLGTLcMPK NKGLFAGDYSETHYSPDGR QENGVNIPcQPQDVKcGR VQSASGITLALFGKWR WRETDLLPR YIEFYVVVDNLMYR YSGNLSTIR	89	SVMP PIII

Table 12 continued.

A.pra.SVMP_21	AAPDIcFNENLKGNSYGYcRQENGVNIP cQPQDVK cGDGKVcSNGEcVDVQTAY EcVMSETLGFEPPYK GNSYGYcRQENGVNIPcQPQDVK KAAPDIcFNENLKGNSYGYcR KRNDNAQLLTGIDFDGPVVGLGYLGTLc MPK NKGLFAGDYSETHYSPDGR QENGVNIPcQPQDVKcGR VQSASGITLALFGEWR YIEFYVVVDNLMYR YSGNLSTIR	80	SVMP PIII
A.pra.SVMP_38	ASAHDcDLAELcSGQSAEcPTDLFQR cGDGKVcSNGEcVDVQTAY DKINVQSASGVTLDLFGEWRETDLLPR GTDYGYcRQENGVSIPcQPQDVK IDFDGNTVGLGYVGTLcKPK INVQSASGVTLDLFGEWRETDLLPR KYIEFYVVVDNLMYR NDNAQLLTR NGHPcENNQGYcYNGAcPTLEK NKGLFAGDYSETHYSPDGR QcIAFFGNDIIVGPDScFDKNLK QENGVSIPcQPQDVK VYEIVNTVNLIYRPLDIHLALIGLEIWSNR DK YIEFYVVVDNLMYR YSAAIVQDHNER	69	SVMP PIII
A.pra.SVMP_36	AAPDIcFNENQKGNSYAYcR ASKHDcDMPELcTGQSAEcPIDLFQR cGDGKVcSNGEcVDVETAY EcVMSATLGFEPPYK GNSYAYcRQENGVNIPcQPK HDcDMPELcTGQSAEcPIDLFQR KAAPDIcFNENQK KRNDNAQLLTGIDFDGQVVGLGYLGTL cMPK NKGLFAGDYSETHYSPDGR QENGVNIPcQPK VQSASGITLALFGKWR WRETDLLPR YIEFYVVVDNLMYRK YSGNLSTIR	68	SVMP PIII

Table 12 continued.

A.pra.SVMP_4	AAPDIcFNENQKGNSYAYcR ASKHDcDMPELcTGQSAEcPIDLFQR cGDGKVcSNGEcVDVETAY EcVMSATLGFEPPYK GNSYAYcRQENGVNIPcQPK HDcDMPELcTGQSAEcPIDLFQR KAAPDIcFNENQK KRNDNAQLLTGIDFDGQVVGLGYLGT LcMPK NKGLFAGDYSETHYSPDGR QENGVNIPcQPK VQSASGITLALFGKWR WRETDLLPR YIEFYVVVDNLMYKK YSGNLSTIR	62	SVMP PIII
A.pra.CRiSP_1	cPATcFcPDK EIVDMHNSFR EKQREIVDMHNSFR MDWYDEAATSAEYWASVcAYDHSPD SSR QREIVDMHNSFRR RSVSPTASNMLK SLAEQSScQDDWIQTK TLNGILcGENIYmSSNPR VDFNSESTR	46	CRiSP
A.pra.SVMP_16	ALNIHVALIGLEIWSNRDK cEDGKVcSNGEcVDVETAY DKINVQSASGVTLDLFGEWRETDLLP R IFEmVNIVNR INVQSASGVTLDLFGEWR LYcTQGNLGGcIVMYYPDTPENGIVET GTKcEDGK NKGLFAGDYSETHYSPDGR	38	SVMP PIII
A.pra.SVMP_39	ASKHDcDLAEHcTGQSADcPTDLFQR cGDGKVcSNGEcVDVQTAY GTDYGYcRQENGVSIPcQPQDVK HDcDLAEHcTGQSADcPTDLFQR INVQSASDVTLALFGEWRETDLLPR NGHPcENNQGYcYNGAcPTLEK QcIAFFGNDIIVGPDScFDKNLK QENGVSIPcQPQDVK TTVFETVNIVNR	29	SVMP PIII
A.pra.SVMP_31	ASKHDcDLAEHcTGQSADcPTDLFQR cGDGKVcSNGEcVDVQTAY GTDYGYcRQENGVSIPcQPQDVK HDcDLAEHcTGQSADcPTDLFQR INVQSASDVTLALFGEWRETDLLPR NGHPcENNQGYcYNGAcPTLEK NKGLFAGDYSETHYSPDGR QcIAFFGNDIIVGPDScFDKNLK QENGVSIPcQPQDVK TTVFETVNIVNR	27	SVMP PIII

Table 12 continued.

A.pra.3FTx_1	AGETVEccKKNEcIVK cLLPMEcPAGENIcYFLWK GcASEcPIPK YEGTLITDWVR	11	3FTx
A.pra.3FTx_7	AGETVEccKKNEcIVK KYDGTSLTDWVK YDGTSLTDWVK	8	3FTx

Abbreviations: 3FTx = three-finger toxin, CRiSP = cysteine-rich secretory protein, and SVMP (PIII) = snake venom metalloproteinase type PIII.

Biological Activity of Crude *Ahaetulla prasina* and *Alsophis portoricensis* Venoms

Both *A. portoricensis* and *A. prasina* venoms lacked phospholipase A_2 activity, which is consistent with the lack of transcripts and proteins for this venom protein superfamily within their venom glands and venom, respectively. Rear-fanged venomous snakes generally lack phospholipase A_2 (PLA_2) activity or exhibit very low activity for this enzyme, with a few exceptions (Hill and Mackessy 2000; Huang and Mackessy 2004; Peichoto et al. 2011a). It appears that PLA_2s within rear-fanged venomous snakes do not serve as vital a role in envenomation as they do for front-fanged venomous snakes, which consistently show high levels of PLA_2 activity and often a diversity of PLA_2s with varying pharmacological effects (Kini 2003; Malhotra et al. 2013).

The venoms also lacked phosphodiesterase (PDE), acetylcholinesterase (AChE), thrombin-like serine proteinase, and kallikrein-like serine proteinase enzyme activities. Very low levels of a PDE transcript were detected within the *A. portoricensis* venom gland transcriptome, but none of these enzymes were detected within the venom proteome. The same situation was observed for *A. prasina* and AChE, low levels of AChE transcripts were expressed within the

venom gland, but either not translated or present at such low concentrations within the venom that the enzyme activity was not detectable. Serine protease transcripts or venom proteins were not detected for either rear-fanged venomous snake. There was an absence of L-amino acid oxidase (LAAO) activity within *A. portoricensis* venom, however, low levels of activity for this enzyme were detected for *A. prasina* (Table 13).

Table 13. Enzyme activity of *Ahaetulla prasina* and *Alsophis portoricensis* venoms.

	Ahaetulla prasina venom	*Alsophis portoricensis* venom
Metalloproteinase activity (ΔA_{342nm}/min/mg)	0.863 ± 0.026	0.960 ± 0.040
L-amino acid oxidase activity (nmol product/min/mg)	0.021 ± 0.005	0

The most prominent enzyme activity was detected using an azocasein substrate. This substrate is degraded by snake venom metalloproteinases (Table 13). This high SVMP activity is not surprising given the large abundances of both SVMP transcripts and proteins within these venom glands and venoms. Venom from *A. portoricensis* has previously been noted for high metalloproteinase activity, and one of the most prominent metalloproteinase from this venom has been isolated and characterized (Weldon and Mackessy 2010; Weldon and Mackessy 2012). Metalloproteinase activity had not been previously explored for *A. prasina,* and the SVMP activity observed for *A. prasina* venom is almost as high as observed for *A. portoricensis* (Table 13).

Rear-fanged snake metalloproteinases have been reported to rapidly degrade fibrinogen subunits (Weldon and Mackessy 2012). Crude *A. prasina* and *A. portoricensis* venom was incubated with human fibrinogen for 60 minutes. Venom from *A. prasina* degraded the alpha subunit of fibrinogen in less than a minute and within five minutes had degraded the beta subunit (Fig. 27). The gamma subunit was not degraded within the 60 minutes. Venom from *A. portoricensis* also degrade the alpha subunit of fibrinogen in less than a minute, consistent with the reported activity of the metalloproteinase Alsophinase within this venom (Weldon and Mackessy 2012). The beta subunit was almost entirely degraded by 60 minutes, while the gamma subunit remained intact (Fig. 27). Surprisingly, even though *A. portoricensis* has been recognized as a snake responsible for human envenomations resulting in localized pain and ecchymosis from metalloproteinase within this venom (García-Gubern et al. 2010) and *A. prasina* venom has received little attention, it appears that SVMPs within *A. prasina* venom have higher fibrinogenolytic activity.

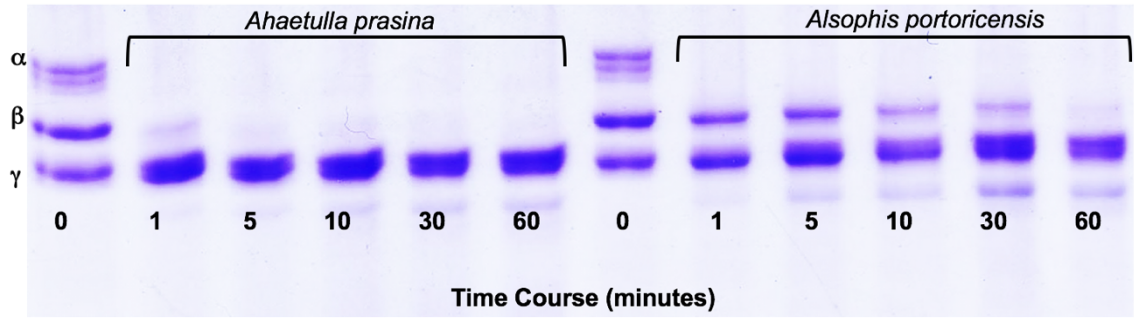


Figure 27. Fibrinogen digest assay with *Ahaetulla prasina* and *Alsophis portoricensis* crude venoms. Crude venoms (20 µg) were incubated with human fibrinogen over a 60-minute time course. Minute 0 was before the addition of any venom and shows all intact fibrinogen subunits. Fibrinogen subunits are labeled α, β, and γ.

Conclusions

Rear-fanged venom gland expression appears to reflect the hemorrhagic/neurotoxic venom dichotomy that characterizes the differences in venom composition between elapids and viperids (McGivern et al. 2014). Both of the venom gland transcriptomes of *A. prasina* and *A. portoricensis* appear to be more viperids-like (McGivern et al. 2014), which have a high abundance of metalloproteinase venom protein transcripts. However, the venom gland transcriptome of *A. prasina* does demonstrates a higher level of expression of 3FTx transcripts than previously observed for viper-like rear-fanged transcriptome expression. The rear-fanged *B. irregularis* and *P. sulphureus* venom gland transcriptomes that have been identified as having elapid-like toxin gene expression with a very high abundance of 3FTx transcripts expressed, both which is still have a larger amount of toxin 3FTx reads than *A. prasina*. The *A. portoricensis* venom gland transcriptome is very similar in overall toxin expression to *Hypsiglena* sp.

There has been several venom gland transcriptome assemblies that have demonstrated that the most abundant venom protein superfamily transcripts correlate with the most abundant venom protein superfamilies within the venom and that post-transcriptional regulation contributes relatively little to the overall venom phenotype (Aird et al. 2013; Rokyta et al. 2015). This is supported within this study, SVMP transcripts were highly abundant within these venom gland transcripts and were also highly abundant within the venom proteomes. However, translation regulation of different isoforms within each venom protein superfamily appears to be possible.

The enzymatic activity exhibited by these crude venoms was also suggestive of high SVMP abundances. This study demonstrates the utility of a combined transcriptomic and proteomic approach to characterize rear-fanged snake venoms, as well as also performing assays for enzyme activity. This approach provides a comprehensive view of rear-fanged snake venom composition and the biological roles the protein components of these venoms provide.

CHAPTER V

FULL-LENGTH VENOM PROTEIN CDNA SEQUENCES FROM VENOM-DERIVED MRNA: EXPLORING COMPOSITIONAL VARIATION AND ADAPTIVE MULTIGENE EVOLUTION

Cassandra M. Modahl and Stephen P. Mackessy*
School of Biological Sciences, University of Northern Colorado, 501 20th St., Greeley, CO 80639-0017, USA

Abstract

Envenomation of humans by snakes is a complex and continuously evolving medical emergency, the treatment of which is made that much more difficult by the diverse biochemical composition of many venoms. In addition, venomous snakes and their venoms provide models for the study of molecular evolutionary processes leading to adaptation and genotype-phenotype relationships. To compare venom protein sequences, venom gland transcriptomes are assembled, which usually requires the sacrifice of snakes for tissue. However, toxin transcripts are also present in venoms, offering the possibility of obtaining cDNA sequences directly from venom. This study provides the first evidence that unknown full-length venom protein transcripts can be obtained from the venoms of multiple species. These unknown venom protein cDNAs are obtained by the use of primers designed from conserved signal peptide sequences within each venom protein superfamily. This technique was used to assemble a partial venom gland transcriptome for the Middle American

Rattlesnake (*Crotalus simus tzabcan*) by amplifying sequences for phospholipases A_2, serine proteases, C-lectins, and metalloproteinases from within venom. Phospholipase A_2 sequences were also recovered from the venoms of several rattlesnakes and an elapid snake (*Pseudechis porphyriacus*), and three-finger toxin sequences were recovered from multiple rear-fanged snake species, demonstrating that the three major clades of advanced snakes (Elapidae, Viperidae, Colubridae) have stable mRNA present in their venoms. These cDNA sequences from venom were then used to explore potential activities derived from protein sequence similarities, evolutionary histories, and positive selection within these large multigene superfamilies. Venom-derived sequences can also be used to aid in characterizing venoms that lack proteomic profiles and identify sequence characteristics indicative of specific envenomation profiles. By requiring only venom to obtain cDNA, this approach provides access to cDNA sequences in the absence of living specimens, even from commercial venom sources, to evaluate important regional differences in venom composition and to study snake venom protein evolution.

Author Summary

This work demonstrates that full-length venom protein messenger RNAs are present in secreted venoms and can be used to acquire full-length protein sequences of toxins from both front-fanged (pit vipers, cobras and relatives) and rear-fanged (colubrid) snake venoms, eliminating the need to use venom glands. Full-length transcripts were obtained from venom samples that were fresh, newly lyophilized, old, field desiccated or commercially prepared, representing a

significant advance over previous attempts which produced only partial sequence transcripts. Transcripts for all major venom protein families (metalloproteinases, serine proteases, C-type lectins, phospholipases A_2 and three-finger toxins) responsible for clinically significant snakebite symptoms were obtained from venoms. These sequences aid in the identification and characterization of venom proteome profiles, allowing for the identification of peptide sequences, specific isoforms, and novel venom proteins. The application of this technique will help to provide venom protein sequences for many snake species, including understudied rear-fanged snakes. Venom protein transcripts offer important insights into potential snakebite envenomation profiles, and the molecular evolution of venom protein multigene families. By requiring only venom to obtain venom protein cDNAs, the approach detailed here will provide access to cDNA-based protein sequences from commercial and other venom sources, facilitating study of snake venom protein composition and evolution.

Introduction

The evolution of venoms among the advanced colubroid snakes has had tremendous adaptive significance and has allowed this clade to diversify rapidly and occupy a diversity of niches globally (Greene 1983). Snake venoms are complex glandular secretions which may contain 2-100+ protein/peptide components with a myriad of biological activities, ranging from potent neurotoxins to rapid-acting myotoxins to hydrolytic enzymes (Mackessy 2010b). These toxins are synthesized and stored in a cephalic venom gland which allows immediate deployment as a chemical weapon, also necessitating intricate storage and

protective mechanisms (Mackessy and Baxter 2006). Venoms likely allowed a transition from mechanical capture and processing of prey to one dependent on chemical means (Kardong and Lavin-Murcio 1993), and during the approximately 100 million year history of snakes (Martill et al. 2015), a diversity of biochemical compositional “strategies” has evolved (Kardong 1996; Jackson 2003). Resulting venom phenotypes can therefore be significantly different, even among closely related species (Mackessy 2010b), and these different phenotypes are often correlated with dietary variables or foraging strategies (Li et al. 2005a; Pawlak et al. 2009; Calvete et al. 2012). Determining detailed venom composition among differing lineages of snakes can provide important connections linking phenotypic variation to specific selective pressures, and linking venom composition to snakebite envenomation effects.

The application of transcriptomic methods has provided insight into venom protein post-transcriptional regulation, as well as documenting isoform diversity and molecular evolutionary trends within large multigene venom protein superfamilies (Harrison et al. 2007; Doley et al. 2008; Durban et al. 2013; Vonk et al. 2013; Casewell et al. 2014). Venom gland transcriptomics has evolved from the generation of ESTs (expressed sequence tags) (Francischetti et al. 2004; Junqueira-de-Azevedo et al. 2006; Zhang et al. 2006; Pahari et al. 2007; Wagstaff et al. 2009; Chatrath et al. 2011; Ching et al. 2012) to more comprehensive next generation sequencing (454 pyrosequencing or Illumina) of total venom gland cDNA (complementary DNA) (Durban et al. 2011; Rokyta et al. 2012b; Aird et al. 2013; Rokyta et al. 2013; Margres et al. 2014; McGivern et al.

2014). However, these methods both currently rely on venom gland tissue to obtain venom protein cDNAs, requiring access to venomous snake tissues and animal euthanasia. The ability to acquire venom protein cDNA sequences from venom has been documented (Chen et al. 2002; Fagundes et al. 2010; Currier et al. 2012; Wang et al. 2013), but this source has not been fully exploited because mRNA yields have been highly variable and very low, and cDNA amplification has not been reliable. Extracellular messenger RNA has been demonstrated to be unusually stable, for at least several years within lyophilized venom (Currier et al. 2012). This alternative source to obtain venom protein cDNAs is a less destructive method because sacrifice of animals is avoided. It also increases the availability of venom protein cDNA sequences to researchers that have limited access to venom gland tissues, as in the case of rare or difficult to acquire snake species, or due to limitations on animal euthanasia protocols. Further, this approach provides the opportunity to generate both venom transcriptomic and proteomic profiles, essentially a genotype-phenotype map, using only the same venom sample from one individual.

The standardized venomics approach of characterizing venoms by separating venom components by HPLC (high performance liquid chromatography), followed by trypsin digestion and tandem mass spectrometry, primarily relies on protein identification from databases such as MASCOT (Calvete 2013). Mass spectral matching to determine peptide sequences is an efficient and less expensive alternative to N-terminal sequencing (Edman degradation). An advantage to this venomic approach is that it is at least semi-

quantitative, allowing inter- and intrapopulational variation in amounts of specific proteins to be estimated. However, many venom proteins, in particular those from rear-fanged venomous snakes, are not present within current databases or are poorly represented, making it difficult to use this methodology to characterize these venoms (Calvete et al. 2012; Calvete 2013; Calvete 2014). The incorporation of transcriptomics into venom proteomics has resulted in venom protein-locus resolution and has been labeled next generation snake venomics (Wagstaff et al. 2009; Aird et al. 2013; Calvete 2014; Margres et al. 2014). Species-specific venom gland transcriptomes aid in the identification and characterization of venom profiles by providing custom databases for tandem mass spectrometry (MS/MS) spectra matching, allowing for the identification of additional peptide sequences, specific isoforms, and novel venom proteins (Ching et al. 2012; Calvete 2014; Margres et al. 2014; McGivern et al. 2014; Chapeaurouge et al. 2015). This study provides support for an approach to obtain species-specific venom protein transcript sequences, including those from rear-fanged venomous snakes, using relatively little starting material (2 mg of lyophilized venom or 100 µl of fresh venom) that does not require venom gland tissue. cDNA derived from this method has great potential to fill gaps within databases and to aid in the characterization of understudied snake venoms.

Acquiring full-length venom protein sequences can also help to identify protein characteristics indicative of serious envenomation profiles, because venom toxins commonly utilize conserved structural folds but produce diverse pharmacological activities. For example, many venoms contain enzymatic

phospholipase A_2s (PLA_2s) of low toxicity (Montecucco et al. 2008; Malhotra et al. 2013); some contain PLA_2s with potent myotoxic activity, and a limited number contain neurotoxic PLA_2s (crotoxin-like complexes or other asparagine-6 containing PLA_2 sequences). Current methods for examining venom protein pharmacological activity can be labor intensive, requiring multiple protein purification steps and functional assays. Phospholipase A_2 functionality based on clustering with other PLA_2s sharing similar sequence has been demonstrated to be a potential *in silico* alternative for predicting specific PLA_2 activity (Malhotra et al. 2013). Knowledge of sequences and sequence similarities to others with noted activity and that are currently promising drug leads can help guide new drug exploration or even help to identify unique characteristics of prey-specific toxins that allow them to specifically target select prey taxa receptors, as in the case of rear-fanged snake venom three-finger toxins (3FTxs).

Rear-fanged snake venoms have not been as well-studied as front-fanged snake venoms, largely due to the difficulties extracting venoms from these snakes and the fact that the large majority of envenomations from these snakes are not life threatening (Chiszar and Smith 2002; Mackessy 2002; Weinstein et al. 2011; Peichoto et al. 2012). However, rear-fanged snake venoms potentially contain proteins that could serve as novel pharmaceutical drug leads or in other applications, such as proteolytic enzymes for protein fragmentation for mass spectrometry (Weldon and Mackessy 2012; Saviola et al. 2014). Rear-fanged snake venoms have also demonstrated unique evolutionary trajectories,

including the presence of the only prey-specific toxins yet identified within snake venoms (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013).

The aim of this study was to obtain abundant venom protein cDNA sequences within venoms to predict protein activities or envenomation symptomology, and to screen for novel sequences that could be of potential biomedical development based on predicted activities and explore the evolutionary trajectories in complex multigene superfamilies. Venom protein evolutionary histories and sites under positive selection were identified for rattlesnake PLA_2s and rear-fanged snake venom 3FTxs. The transcripts obtained provided the opportunity to compare patterns of synonymous and nonsynonymous changes in protein coding sequences with functional and adaptive significance for these large multigene superfamilies. From the rattlesnake PLA_2 transcripts identified, sequence characteristics shared with known viper neurotoxic PLA_2s were observed, demonstrating the utility of using these methods to obtain predictive venom activities and envenomation profiles.

Materials and Methods

Snake Venoms and Reagents

Many of the snake venoms were collected manually from venomous snakes maintained in the University of Northern Colorado Animal Resource Facility in accordance with UNC-IACUC protocols (#9204 and 1302D-SM-S-16) and collecting permits from state and federal agencies (Arizona Game and Fish Department #MCKSY000221 and #SP727017; Colorado Parks and Wildlife #15HP974; U.S. Fish and Wildlife Service #MA022452-0). Venom was collected

from front-fanged vipers *Crotalus scutulatus scutulatus* (Mohave Rattlesnake; SE Arizona)*, Crotalus cerastes* (Sidewinder; S Arizona), *Crotalus oreganus cerberus* (Arizona Black Rattlesnake; E Arizona)*, Crotalus oreganus concolor* (Midget Faded Rattlesnake; S. Wyoming)*,* and *Sistrurus miliarius barbouri* (Florida Pigmy Rattlesnake; central Florida) by placing an RNase Away (Thermo Fisher Scientific Inc., U.S.A.)-treated 100 µl capillary tube over each fang and gently massaging the gland; 100 µl of venom was then immediately added to 1 mL of TRIzol (Life Technologies, CA, U.S.A.). Venom from *Crotalus simus tzabcan* (Middle American Rattlesnake; Yucatán Peninsula, México) was extracted into a sterile beaker and 25 µl, 50 µl and 100 µl of venom were each immediately added to 1 mL of TRIzol; the remaining venom was then centrifuged (9500 x g for 5 minutes), lyophilized and stored at -20 °C until used. Venom from *Crotalus molossus nigrescens* (Mexican Black-tailed Rattlesnake; Morelia, México) was collected in the field, desiccated, and stored at -20 °C until used. Lyophilized venom from *Crotalus pricei* (Twin-spotted Rattlesnake; SE Arizona) was collected from a captive snake and stored frozen (-20 °C) with desiccant for approximately 20 years; lyophilized *Crotalus basiliscus* (Mexican West Coast Rattlesnake; W. México) venom was purchased from the Miami Serpentarium (Lot#CB15SZ) and lyophilized venom from an elapid snake, *Pseudechis porphyriacus* (Red-bellied Black Snake; E Australia), was a gift from Venom Supplies Pty Ltd (Tanunda, South Australia).

Venoms from the rear-fanged snakes *Boiga irregularis* (Brown Treesnake, Guam)*, Boiga dendrophila* (Mangrove Snake; Indonesia)*, Boiga nigriceps* (Black-

headed Catsnake; Indonesia), *Trimorphodon biscutatus lambda* (Sonoran Lyre Snake; Portal, AZ), and *Alsophis portoricensis* (Puerto Rican Racer; Guana Island, British Virgin Islands) were extracted using the method of Hill and Mackessy (1997) with subcutaneous injections of ketamine-HCl (20-30 mg/kg) followed by pilocarpine-HCl (6 mg/kg). Venom was collected by placing RNase Away-treated 100 µl capillary tubes over each enlarged rear maxillary tooth (Hill and Mackessy 1997), and venom (100 µl) was then added to 1 mL of TRIzol. For the rear-fanged snakes *Boiga cynodon* (Dog-toothed Catsnake; Indonesia), *Oxybelis fulgidus* (Green Vine Snake; Central America), and *Ahaetulla prasina* (Asian Vine Snake; Indonesia), venom was collected using the same protocol without RNase Away treated capillary tubes, centrifuged (9500 x g for 5 minutes), lyophilized, and stored at -20 °C until used.

The 3' RACE System kit, PCR SuperMix High Fidelity polymerase, custom oligonucleotides, DNase I, and *Escherichia coli* DH5 α competent cells were purchased from Life Technologies, CA, U.S.A. The plasmid Quick Clean 5M Miniprep Kit was from GenScript, Inc (Piscataway Township, NJ, U.S.A), and the pGEM-T Easy Vector System and Wizard SV gel and PCR clean-up system from Promega, Inc. (Madison, WI, U.S.A.) All other reagents were purchased from Sigma (St. Louis, MO, U.S.A).

Venom RNA Isolation and cDNA Synthesis

RNA was purified from 2 mg of lyophilized venom or 100 µl of freshly collected venom that had been added to 1 mL of TRIzol following the recommended TRIzol RNA protocol: after incubation for 5 minutes, 200 µL of

chloroform was added to each tube, tubes were centrifuged at 12,000 x g for 15 minutes, aqueous upper phases were transferred to new RNase-free tubes, and 500 µL 100% isopropanol added to each aqueous phase to precipitate RNA. Tubes were incubated at room temperature for 10 minutes and centrifuged at 12,000 x g for 10 minutes. Supernatant was removed and the resulting RNA pellet (not visible) washed with 1 mL 75% ethanol. Another centrifuge step at 7,500 x g for 5 minutes was performed and supernatant poured off. An -20 °C overnight incubation in 300 µL 100% ethanol with 40 µL 3 M sodium acetate was then performed to increased RNA yields, and the following day tubes were centrifuged at 10,000 x g for 15 minutes, supernatant removed, and total RNA resuspended in 16 µL nuclease-free H_2O. To evaluate the effect of different amounts of lyophilized and fresh venom on cDNA yields, 5 mg, 10 mg, and 20 mg of lyophilized venom or 10, 25, 50 or 100 µL of crude fresh venom from *C. s. tzabcan* was added to TRIzol reagent and processed as above. For rear-fanged snake venoms, extraction methods resulted in retention of significant amounts of contaminating DNA; therefore, an Amplification Grade DNase I digestion after RNA isolation was performed at room temperature for 15 minutes, followed by the addition of 1 µl 25 mM EDTA (pH 8.0) and a 15 minute 65 °C incubation. cDNA synthesis from total RNA was accomplished using the 3’ RACE System following the manufacturer’s protocols. The oligo(dT) adaptor primer provided with the kit initiated reverse transcriptase cDNA synthesis and effectively selected for polyadenylated mRNAs.

Rapid Amplification of cDNA Ends (3'RACE)

Sense primer sequences were designed from conserved signal peptide regions for each venom protein superfamily (Table 14). To identify conserved signal peptide sequences, multiple sequence alignments within MEGA v6.06 (Tamura et al. 2013) using MUSCLE (Edgar 2004) were performed for each venom protein superfamily, with representative sequences obtained from the NCBI (National Center for Biotechnology Information) nucleotide database. Each sense primer was used in a reaction with the 3'RACE system AUAP antisense primer 5'-GGCCACGCGTCGACTAGTAC-3'. For Mojave toxin, published sense and antisense primers were used for both acidic and basic subunits (Wooldridge et al. 2001). Twenty-three µL of PCR SuperMix High Fidelity polymerase was used with 1-2 µL of cDNA template and 0.5 µL of each primer (sense and antisense). PCR was performed with seven touchdown cycles of 94°C for 25 seconds, 52°C for 30 seconds, and 68°C for two minutes. Thirty additional cycles followed with 94°C for 25 seconds, 48°C for 30 seconds, and 68°C for two minutes with a final 68°C extension for five minutes. The amplified products were observed on a 1% agarose gel, and bands of the estimated transcript sizes (based on previous published transcripts) were excised and then purified using the Wizard SV gel and PCR clean-up system.

Table 14. List of primers used for amplification of transcripts within specific venom protein families.

Venom Protein	Primer sequence	Reference
Rattlesnake phospholipase A_2	5'-GTCTGGATTCRGGAGGATGAGG-3'	Current study
Elapid phospholipase A_2	5'-CTGYTGBTGANBKTG-3'	Current study
Rattlesnake metalloproteinase (PII and PIII)	5'-AATCYAGSCTCCAAAATGATC-3'	Current study
Rear-fanged snake metalloproteinase	5'-ATGATCCAAGYTCTCTTGRTWACTATATRCTTAG-3'	Current study
Rattlesnake serine protease	5'-ATGGTGCTGATCAGAGTGCTAGCAAACCTTCT-3'	Current study
Rattlesnake C-type lectin	5'-ATGKGGCRATTSAYC-3'	Current study
Mojave toxin subunit A	Sense: 5'-GGTATTTCGTACTACAGCTCTTACGGA-3' Antisense: 5'-TGATTCCCCCTGGCAATT-3'	Wooldridge et al., 2001
Mojave toxin subunit B	Sense: 5'-AACGCTATTCCCTTCTATGCCTTTTAC-3' Antisense: 5'CCTGTCGCACTCACAAATCTGTTCC-3'	Wooldridge et al., 2001

Cloning and Sequencing of Venom cDNA

Amplified and purified cDNA was ligated into the pGEM-T Easy Vector System and transformed into *Escherichia coli* DH5 α competent cells following the manufacture's recommended protocol. Transformed *E. coli* were grown on nutrient rich agar plates overnight at 37 °C with ampicillin, IPTG and β-galactosidase for white/blue colony selection. Recombinant plasmids were selected from agar plates, and *E. coli* colonies picked for viper PLA_2 sequences

were as follows: three colonies were picked for *C. cerastes*, three for *S. m. barbouri*, four for *C. m. nigrescens,* six for *C. o. cerberus,* six for *C. basiliscus,* eight for *C. pricei,* twelve for *C. o. concolor,* and fifteen were picked for *C. s. tzabcan.* In addition to the PLA_2 sequences, eight colonies for serine proteases, ten colonies for C-type lectins, and eighteen colonies for metalloproteinases were chosen from *C. s. tzabcan* venom to obtain transcripts for all major venom protein families. For rear-fanged colubrid snake 3FTxs, the following number of *E. coli* colonies were selected: three were picked for *T. b. lambda,* three for *A. prasina,* six for *O. fulgidus,* four for *B. nigriceps,* ten for *B. cynodon,* twenty for *B. dendrophila,* and nineteen were picked for *B. irregularis.* A sampling of 3FTxs from *Boiga sp.* was chosen because of the three currently identified prey-specific 3FTxs, two have been found in *Boiga* species (Pawlak et al. 2006; Pawlak et al. 2009). Three-finger toxin sequences from rear-fanged snakes, especially those from *Boiga sp.* and *O. fulgidus*, can provide insight into the evolution of 3FTx prey-specific binding affinities. Three colonies were also picked for metalloproteinase transcripts in *A. portoricensis* venom. For the elapid snake *P. porphyriacus*, five colonies from amplified PLA_2s were picked. The numbers of colonies picked varied depending on the number of expected isoforms within each snake venom protein family and also because some primers were still being evaluated for specificity (only a few colonies were selected in these cases). Each *E. coli* colony was placed into 2 mL LB broth with 1 µL/mL ampicillin, and shaken overnight at 37 °C. Plasmid copies for each *E. coli* colony were than purified using the Quick Clean 5M Miniprep Kit and were sequenced at the DNASU

facility (Arizona State University, AZ, U.S.A) using Big Dye V3.1 chemistry with samples processed on an Applied Biosystems 3730XL Sequence Analysis Instrument.

Sequence Analysis

Sequences were viewed with 4Peaks software (http://nucleobytes.com/index.php/4peaks) and base pairs with acceptable quality scores (Phred score >20) were retained for analysis. Redundant sequences were removed. Sequences were identified with BLASTx (Basic Local Alignment Search Tool) on the NCBI server, limiting the search to “Serpentes (taxid: 8570)” proteins. Protein identities were considered significant if they fell below an e-value threshold of e^{-4} and shared sequence similarity to other known snake venom proteins. Sequences were translated to their corresponding amino acid sequence and trimmed in MEGA v6.06 (Tamura et al. 2013), then aligned with MUSCLE (Edgar 2004) and manually checked.

Phylogenetic analysis was completed with MrBayes v3.2.4 (Huelsenbeck and Ronquist 2001) using models selected by PartitionFinder v1.1 (Lanfear et al. 2012). PartitionFinder v.1.1 models selected were favored using Akaike Information Criterion. These datasets were then run in duplicate using MrBayes v3.2.4 with the default of three heated and one cold chain for $1x10^7$ generations, sampling every 1,000 generations, and with the first 10% discarded as burn-in. Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) was used to check for run convergences. Consensus tree figures were prepared with FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Sites under positive selection within the rattlesnake PLA_2 and rear-fanged snake 3FTx superfamilies were determined from likelihood-ratio tests with the use of the PAML (Phylogenetic Analysis by Maximum Likelihood) version 4.7 codeml program (Yang 2007). Variable selective pressures among sites and branches were used, incorporating the results of the Bayesian analysis above as the user defined tree. Likelihood ratio tests for positive selection were achieved by comparing models M0, M1, M2, M7 and M8. This included comparing M1 (a model of neutral evolution) versus M2 (a model of positive selection), and comparing M7 (beta) versus M8 (beta and ω). To test for the model with best fit, twice the negative difference in log likelihoods between the models were compared to a X^2 distribution with 2 degrees of freedom with a Bonferroni correction to determine critical *P* values and significance. To estimate overall *dN/dS,* the M0 model fitting a single ratio for all sites and branches (averaging across sites) was used. The M0 model is not the best indicator of positive selection, since only some sites are under positive selection, but it does give an overall positive selection rate across coding sequences. Bayes Empirical Bayes (BEB) was used to identify amino acids under positive selection by calculating posterior probabilities, sites with posterior probabilities greater than 99% were reported.

Models were predicted with the Phyre2 webserver (Kelley et al. 2015). Within Phyre2, the program SuSPect (Yates et al. 2014) was used to identify sites sensitive to mutation and to generate colored models. SuSPect predictions were checked for agreement with both conservation predictions within Phyre2

using the information-theoretic approach based on Jensen-Shannon divergence (Capra and Singh 2007) and codeml predicted sites evolving under positive selection.

Results

Venom RNA Isolation

Venom RNA concentrations were determined using both Nanodrop 2000 (Thermo Fisher Scientific, NY, U.S.A) and Qubit 2.0 Fluorometer (Life Technologies, CA, U.S.A) with a high sensitivity RNA assay kit. Various RNA isolation kits and reagents were used to determine which method produced the greatest RNA yields and amplification success (Table 15). The TRIzol RNA isolation protocol described in the methods was found to produce the most consistent results when isolating venom RNA and amplifying transcripts; however, the Direct-zol RNA kit (Zymo Research, CA, U.S.A) and *mir*Vana miRNA isolation kit (Life Technologies, CA, U.S.A) were also successful. Dynabeads from Life Technologies, CA, U.S.A were tried using a previously published technique for isolating extracellular mRNA within venom (Chen et al. 2002; Currier et al. 2012), as well as a FastTrack MAG mRNA isolation kit (Life Technologies, CA, U.S.A) and a RNeasy mini kit (QIAGEN, CA, U.S.A), but cDNA amplification did not produce visible PCR products (Table 15).

Table 15. RNA and mRNA isolation protocols used to obtain extracellular RNA within venom, and the resulting yields and cDNA amplification success.

RNA or mRNA isolation procedure	Total RNA or mRNA yields (Nanodrop)	Total RNA or mRNA yields (Qubit)	Successful cDNA amplification
TRIzol	1.10-13.60 µg (RNA)	< 20 ng/µl (RNA)	+
*mir*Vana miRNA isolation kit	0.33-1 µg (RNA)	< 20 ng/µl (RNA)	+
Direct-zol RNA kit	0.08-0.13 µg (RNA)	nd	+
RNeasy mini kit	0.08-0.30 µg (RNA)	nd	−
Dynabeads	0.20-1.19 µg (mRNA)	nd	−
FastTrack MAG mRNA isolation kit	0.01-0.03 µg (mRNA)	nd	−

nd = not determined; + = successful amplification, - = unsuccessful.

The TRIzol reagent Nanodrop readings for RNA yields within rattlesnake venom varied from 69 ng/µl (1.1 µg of total RNA isolated from 2 mg of lyophilized rattlesnake venom) to 683.3 ng/µl (10.9 µg of total RNA isolated from 100 µl of fresh rattlesnake venom) (Table 15). Rear-fanged snake venoms consistently showed slightly higher yields (10.3 µg – 13.6 µg of total RNA) as determined by Nanodrop; however, this appeared to be mostly due to the 260 nm readings of contaminating DNA, and all RNA isolated from rear-fanged venomous snakes required a DNase I digestion before PCR to prevent nonspecific amplification. When fresh venom was used, 100 µl of *C. s. tzabcan* venom yielded a total RNA amount (Nanodrop) of 6.1 µg, 50 µl yielded 8 µg, and 25 µl yielded 10.4 µg.

However, when the same volume amounts were used for cDNA synthesis and amplification, successful amplification of PLA_2 transcripts tended to decrease with decreasing venom input (Fig. 28A). When different lyophilized venom amounts were used, 20 mg yielded 6.1 µg, 10 mg yielded 5.2 µg, 5 mg yielded 6.5 µg, 2 mg yielded 5.3 µg, and 1 mg yielded 8.4 µg of total RNA (Nanodrop). As seen with the total RNA amounts from fresh venom, the Nanodrop amount of total RNA from lyophilized venom from did not demonstrate a clear relationship to PLA_2 transcript amplification success, and 2 mg produced the highest concentration of PLA_2 amplicon (Fig. 28B). All Nanodrop readings did show low 260/280 and 260/230 ratios, indicating low purity. The typical 260/280 ratio observed was 1.5 and 1.6 for 260/230. Qubit results revealed values less than 20 ng/µl, below instrument detection, for all measured samples (Table 15).

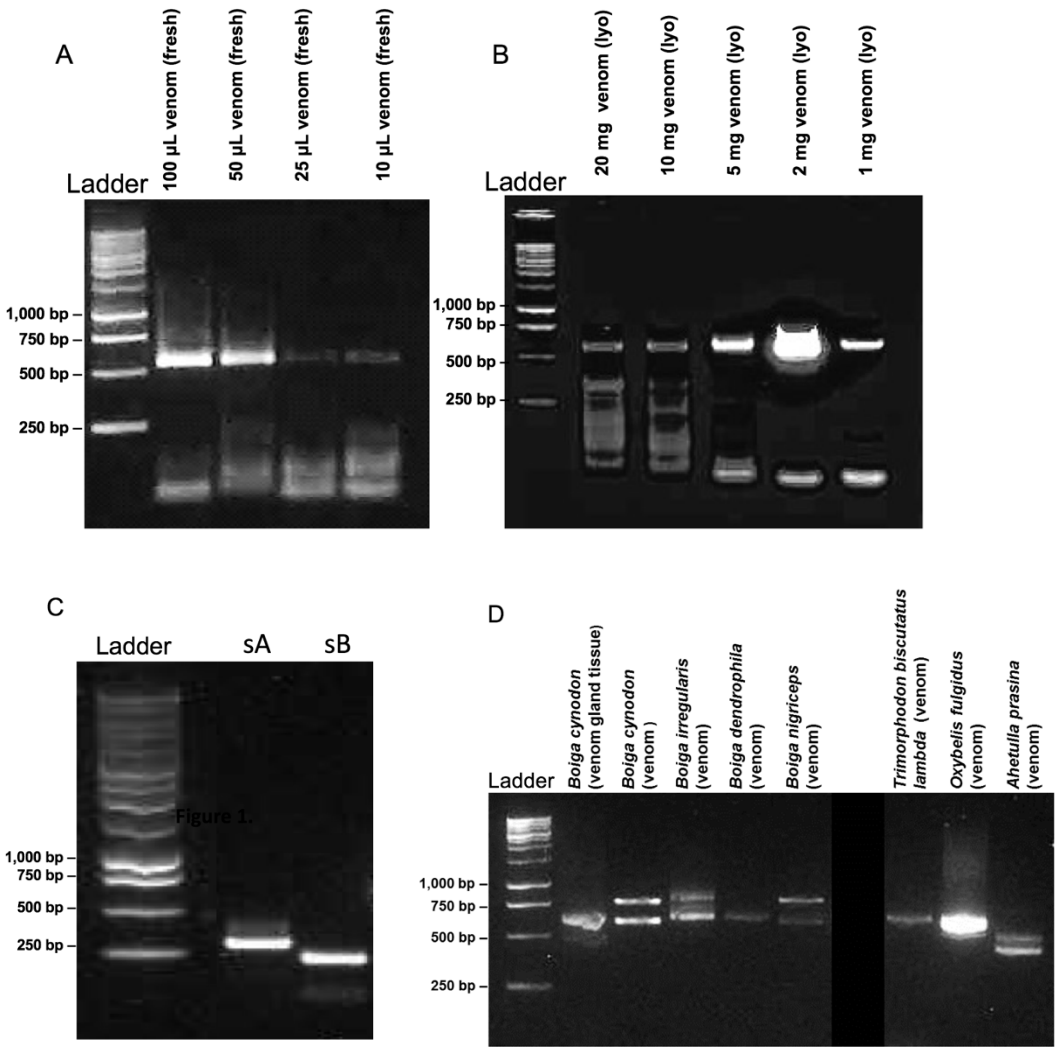


Figure 28. Agarose gel electrophoresis showing cDNA transcripts amplified from mRNA in venomous snake venoms. Various amounts of fresh venom (A) and lyophilized venom (B) from *Crotalus simus tzabcan* were used to determine optimal amplification conditions for phospholipase A_2 transcripts. Regions of phospholipase A_2 Mojave toxin subunit A (sA) and subunit B (sB) cDNAs were amplified to demonstrate transcript detectability in *Crotalus scutulatus scutulatus* venom (C). Three-finger toxin cDNA sequences were also amplified from mRNA derived from venoms of rear-fanged snakes (D).

It was possible to isolate RNA from both fresh venom and from lyophilized venom that had been stored at -20°C (including after 20 years of storage), as well as from venom desiccated in the field and venom purchased from a commercial venom supply source. Both front-fanged and rear-fanged venomous snakes were found to have extracellular RNA within their venoms; this is the first report of mRNA in the venom of rear-fanged venomous snakes, and RNA was isolated from both freshly collected and lyophilized rear-fanged snake venoms.

Venom Protein cDNA Sequences within Venom

As proof of concept, the presence of cDNA amplicons from *C. s. scutulatus* venom, obtained using published primer sequences (Wooldridge et al. 2001), was used to confirm presence of the two Mojave toxin subunits (acidic and basic chains; Fig. 28C). These basic and acidic subunits were both sequenced and found to be 100% identical to the published sequences for *C. s. scutulatus* (PA2A_CROSS and PA2Ba_CROSS) (Aird et al. 1990; Bieber et al. 1990), demonstrating that cDNAs of mRNA within venom can be used to detect the presence of specific expressed venom protein transcripts. In this case, the presence and abundance of crotoxin/Mojave toxin-like acidic and basic subunits is strongly indicative of neurotoxic envenomation symptoms characteristic of human envenomations by these rattlesnakes.

3'RACE with sense primers designed from conserved sequences of the signal peptide or the 5'UTR (untranslated region) of transcripts (Table 14) was used to amplify cDNAs for a diversity of PLA_2s, metalloproteinases, serine

proteases, C-type lectins, and 3FTxs from viperid, elapid and rear-fanged snake venoms. This is the first time that cDNA derived from venom transcripts has been used to obtain unknown sequences for such a diversity of venom protein families. For Middle American Rattlesnake venom (*C. s. tzabcan),* full-length cDNA sequences were successfully amplified for the major venom proteins present within this rattlesnake's venom (Castro et al. 2013) in spite of limited colony sampling (Table 16). A partial venom gland transcriptome was assembled, focusing on venom protein transcripts that significantly contribute to envenomation symptomology, including metalloproteinases, serine proteases, and C-type lectins. There are likely many more unique C-type lectins, serine proteases, and metalloproteinase transcripts within *C. s. tzabcan* venom, but the intent here was to demonstrate the presence of diverse, intact venom protein transcripts. Greater diversity of C-type lectins have been identified within other rattlesnake venom gland transcriptome assemblies using next generation sequencing (Rokyta et al. 2012b; Aird et al. 2013; Rokyta et al. 2013), but the vast majority of these C-type lectins transcripts were found to be present in very low abundance.

Table 16. Venom protein transcripts amplified from the venom of the Middle American Rattlesnake (*Crotalus simus tzabcan*) for the four dominant protein families present.

Venom Protein Family	# Colony Picks	# Unique Isoforms
Metalloproteinase (PII and PIII)	18	4
Phospholipase A_2	15	4
C-type Lectin	10	4
Serine Protease	8	4

The diversity of PLA_2 isoforms appeared to vary for each rattlesnake species (Table 17). For *C. pricei,* only one unique PLA_2 sequence was discovered in eight selected clones, while *C. m. nigrescens* had three unique sequences found in the selection of only four clones. This demonstrates that considerable variation in the number of rattlesnake PLA_2 isoforms for each species exists, with no clear positive trend between the number of colonies sequenced and an increase in number of unique sequences (Pearson's correlation test; df = 6, r = 0.6684, p = 0.0699); instead, isoform diversity appeared to be dependent on the species. However, an increase in the number of clones sequenced should increase the chance of observing less abundant isoforms (Pahari et al. 2007; Cardoso et al. 2010).

Table 17. Relationship between number of colony picks and number of observed unique phospholipase A_2 isoforms.

Species	# Colony Picks	# Unique PLA_2 Isoforms
Crotalus simus tzabcan	15	4
Crotalus oreganus concolor	12	3
Crotalus pricei pricei	8	1
Crotalus oreganus cerberus	6	2
Crotalus basiliscus	6	3
Crotalus molossus nigrescens	4	3
Crotalus cerastes cercobombus	3	1
Sistrurus miliarius barbouri	3	1
Pseudechis porphyriacus	5	2

Of the fifteen PLA_2 clones selected from *C. s. tzabcan,* two unique sequences were similar to sequences from crotoxin or Mojave toxin-like acidic A chain (Fig. 29). One sequence (C_s_tzabcan1) was the most abundant, with six identical clones, and it was 99% identical in amino acid sequence to crotoxin acidic A chain (PAIA_CRODU) from the Cascabel Rattlesnake (*Crotalus durissus terrificus*), while the second sequence (C_s_tzabcan4) had only one clone and was 88% identical to the Mojave toxin acidic (A) chain (PA2A_CROSS) from the Mojave Rattlesnake (*C. s. scutulatus*). This less abundant acidic chain sequence revealed that isoform variation within the acidic (A) chains of the toxin exists for *C. s. tzabcan*.

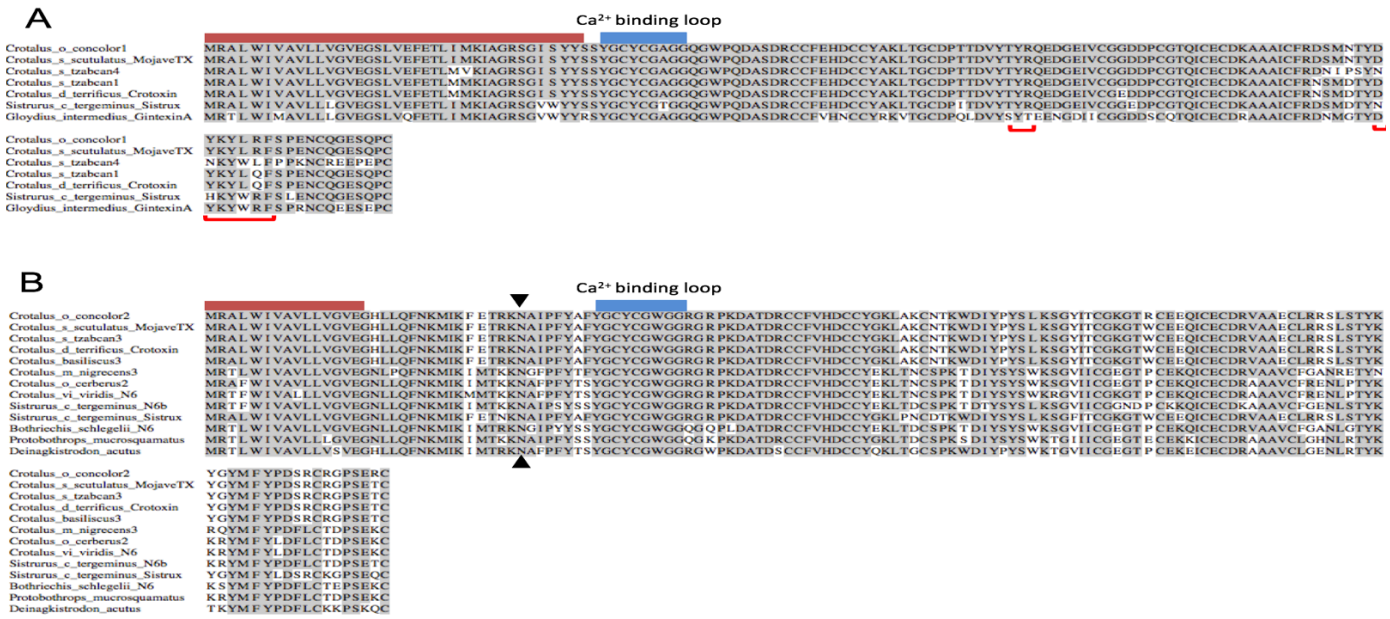


Figure 29. Aligned Group IIA phospholipase A_2 acidic (A) and basic (B) subunit isoforms. A) Sequence alignments of the acidic (A) subunit of crotoxin or Mojave toxin homologs with identical residues shaded; the conserved signal peptide region is indicated by the red bar, and the Ca^+ binding loop is indicated by the blue bar. Regions of the A subunit which are post-translationally cleaved in the mature protein are indicated by red brackets below sequences. GenBank accession numbers of known toxins are as follows: Crotalus_s_scutulatus_MojaveTX (U01026.1), Crotalus_d_terrificus_CrotoxinA (X12606.1), Sistrurus_c_tergeminus_SistruxinA (Q6EAN6.1), and Gloydius_intermedius_GintexinA (AID56658.1). B) Sequence alignments of the basic (B) subunit of crotoxin/Mojave toxin homologs with identical residues shaded; the conserved signal peptide region is indicated by the red bar, and the Ca+ binding loop is indicated by the blue bar. The asparagine-6 (N6) associated with neurotoxic PLA_2 functionality is indicated by arrowheads. GenBank accession numbers of known toxins are as follows: Crotalus_s_scutulatus_MojaveTX (U01027.1), Crotalus_d_terrificus_CrotoxinB (X12603.1), Crotalus_v_viridis_N6 (AF403138.1), Sistrurus_c_tergeminus_N6 (AY355169.1), Sistrurus_c_tergeminus_SistruxinB (Q6EER2.1), Bothriechis_schlegelii_N6 (AY355168.1), Protobothrops_mucrosquamatus (AF408409.1), and Deinagkistrodon_acutus (X77649.1

A crotoxin-like basic (B) chain was also sequenced from *C. s. tzabcan* venom, with three clones that were 100% identical in amino acid sequence to crotoxin subunit CBc from *C. d. terrificus* (PA2BC_CRODU), providing molecular evidence that *C. s. tzabcan* from this study has an abundance of available PLA_2 transcripts to form a neurotoxic complex similar to that of *C. d. terrificus* (Fig. 29). The PLA_2 cDNAs obtained from *C. basiliscus* also had multiple clones containing crotoxin-like basic B chain sequences (C_basiliscus3), which were 100% identical in amino acid sequence to crotoxin basic subunit CBc of *C. d. terrificus* (PA2BC_CRODU) (Fig. 29B). However, crotoxin-like acidic A chain sequences were not discovered, either due to a lack of sufficient sampling or absence from this venom sample. Crotoxin-like protein complexes have been previously observed in *C. basiliscus* venom (Chen et al. 2004).

Mojave toxin-like PLA_2 sequences were obtained from *C. o. concolor* venom (Fig. 29), which has been previously recognized as containing a neurotoxic PLA_2 complex (concolor toxin) similar to Mojave toxin, identified from Ouchterlony immunodiffusion, immunoelectrophoresis, ELISA, and Western blot analyses (Glenn and Straight 1977; Weinstein et al. 1985); however, the full sequence has not been published. Concolor toxin acidic (A) chain was found to share 100% sequence identity with Mojave toxin acidic (A) chain (PA2A_CROSS) from *C. s. scutulatus;* however, the concolor basic (B) subunit was found to be more similar to crotoxin subunit CBc (PA2BC_CRODU) from *C. d. terrificus,* sharing 99% amino acid sequence identify with this crotoxin basic (B) chain.

Interestingly, *C. o. cerberus* and *C. m. nigrescens* venoms also had PLA_2 sequences that contained the asparagine-6 (N6) substitutions associated with myotoxic/neurotoxic activity, also a feature of the basic (B) subunits of crotoxin and Mojave toxin (Fig. 29). All other PLA_2 sequences were similar in sequence to acidic PLA_2s that show edema-inducing activity and myotoxicity, which corresponds to the envenomation symptomology seen in bites from these species. The *S. m. barbouri* PLA_2 sequence reported here was found to be 100% identical to the amino acid sequence of a previously reported PLA_2 from *S. m. barbouri* venom (ABY77929.1 (Gibbs and Rossiter 2008).

There were only two unique PLA_2 transcripts identified in five colonies selected from elapid (*P. porphyriacus*) venom (Table 17). However, of the identified PLA_2 sequences, 4/5 clones were 100% identical in mature protein sequence (P. porphyriacus1) to a previously identified PLA-1 precursor (AAZ22667.1) from *P. porphyriacus* venom gland (Pierre et al. 2005), and the other unique isoform (P. porphyriacus2) was 99% identical to PLA_2 pseudexin B chain (PA2BB_PSEPO), also from *P. porphyriacus* (Schmidt and Middlebrook). Again, sequences determined from venom-derived mRNAs are identical to previously reported venom protein sequences (Fig. 30), validating this method.

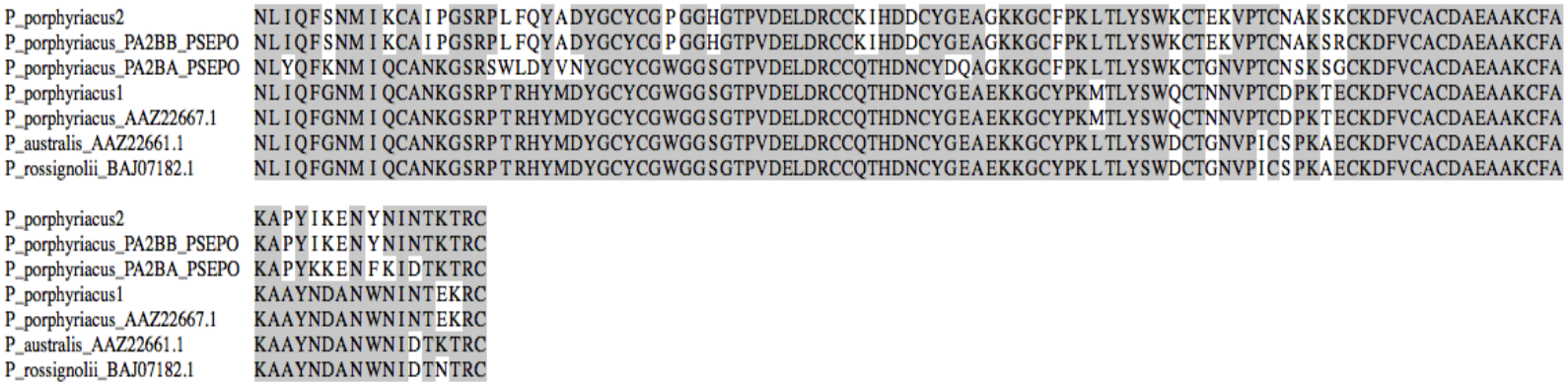


Figure 30. Aligned *Pseudechis* Group IA phospholipase A_2 (PLA_2) mature protein sequences. Amplified PLA_2 sequences from *Pseudechis porphyriacus* venom were aligned to known PLA_2 sequences from the same species and known PLA_2 sequences from two other *Pseudechis* species. Sequences from *P. porphyriacus* venom showed 99% and 100% identity to *P. porphyriacus* mature PLA_2 protein sequences that had been obtained from venom gland tissue or via Edman degradation.

Full-length venom protein transcripts were also identified from rear-fanged snake venoms. Thirty full-length 3FTx sequences were obtained using a degenerate sense primer designed from multiple sequence alignments with published non-conventional 3FTx sequences (Fig. 31). Three-finger toxin transcripts were found in the venoms of *T. b. lambda, A. prasina, O. fulgidus, B. nigriceps, B. cynodon, B. dendrophila,* and *B. irregularis* (Fig. 31).

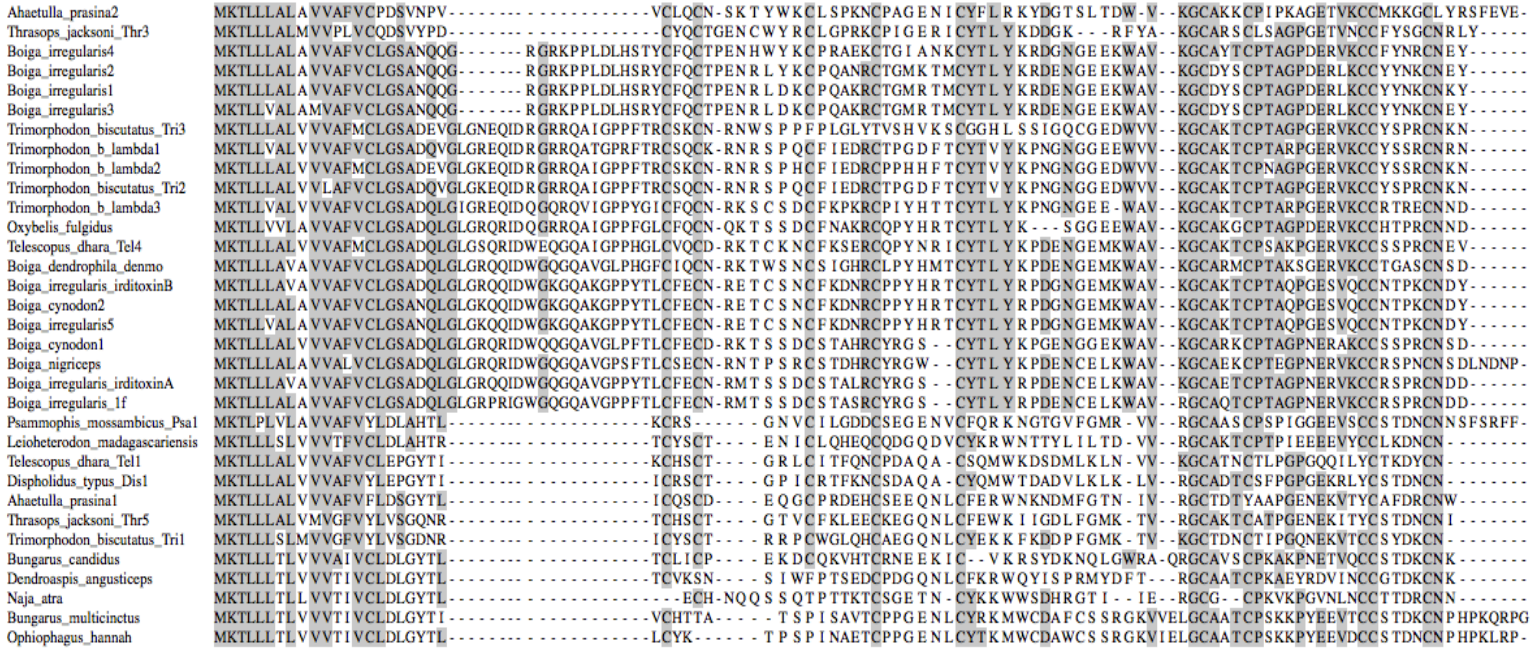


Figure 31. Aligned sequences of non-conventional three-finger toxins (3FTxs) from rear-fanged and Elapid venomous snakes. Venom derived 3FTx sequences obtained from *Boiga irregularis, B. dendrophila, B. nigriceps, B. cynodon, Oxybelis fulgidus, Ahaetulla prasina,* and *Trimorphodon biscutatus lambda* were aligned with various other rear-fanged and Elapidae species with identical nucleotide sequences shaded. GenBank accession numbers are as follows: Trimorphodon_biscutatus_Tri3 (EU029678.1), Trimorphodon_biscutatus_Tri2 (EU029677.1), Telescopus_dhara_Tel4 (EU029686.1), Boiga_dendrophila_denmo (DQ366293.1), Boiga_irregularis_irditoxinB (DQ304539.1), Boiga_irregularis_irditoxinA (DQ304538.1), Boiga_irregularis_1f (GBSH01000015.1), Thrasop_jacksoni_Thr3 (EU029685.1), Dispholidus_typus_Dis1 (EU029674.1), Telescopus_dhara_Tel1 (EU029675.1), Thrasops_jacksoni_Thr5 (EU036635.1), Trimorphodon_biscutatus_Tri1 (EU029675.1), Naja_atra (AF031472.1), Bungarus_multicinctus (AF056400.1), Ophiophagus_hannah (FJ952515.1), Psammophis_mossambicus_Psa1 (EU029669.1), Leioheterodon_madagascariensis (EU029676.1), Bungarus_candidus (AY057878.1), and Dendroaspis_angusticeps (AF241871.1).

Although none of the three 3FTx sequences from *T. b. lambda* venom were 100% identical to previously published 3FTx sequences from *T. b. lambda* (Fry et al. 2012), they did cluster with the previous *T. b. lambda* sequences within a well-supported clade also containing other 3FTx sequences from New World rear-fanged venomous snakes (Fig. 32). A greater diversity of 3FTxs within *T. b. lambda* venom is possible considering that only three clones were picked for this study and all were unique 3FTx sequences; differences observed could also be due to locality-specific transcript variation. Two unique 3FTx sequences were found in *A. prasina* venom and only one unique 3FTx sequence in *O. fulgidus.* The sequence from *O. fulgidus* venom was not identical to the previously characterized fulgimotoxin, which was based on N-terminal Edman degradation sequencing (Heyborne and Mackessy 2013); however, it did show 95% amino acid sequence identity.

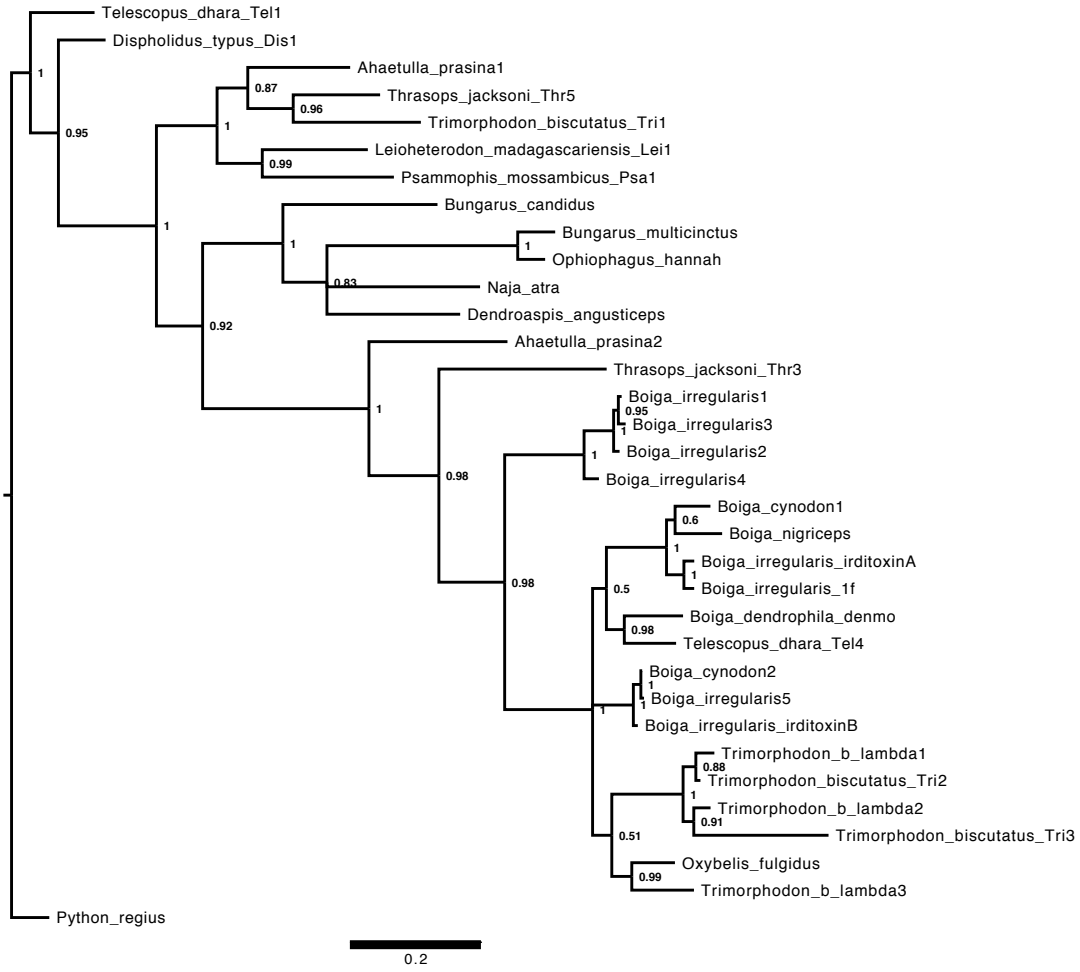


Figure 32. Bayesian sequence similarity tree depicting non-conventional three-finger toxin (3FTx) relationships. 3FTx cDNA sequences were obtained from *Boiga irregularis, B. dendrophila, B. nigriceps, B. cynodon, Oxybelis fulgidus, Ahaetulla prasina* and *Trimorphodon biscutatus lambda,* and sequence relationships with various known rear-fanged and Elapidae 3FTxs are shown. Posterior probabilities are shown for each node, and GenBank accession numbers correspond to sequences from Fig. 31.

Only two unique 3FTx sequences were found in the venom of *B. cynodon* (from ten selected colonies) and one in *B. nigriceps* venom (from four selected colonies). Many of the clones sequenced from rear-fanged snake venoms were of poor quality and were culled, so more sequences are likely present. Fifteen unique 3FTx sequences were revealed in *B. dendrophila* venom, but the majority

were missing complete signal peptide sequences and therefore were omitted from further analysis because it is unknown if these transcripts produce proteins that are secreted in the venom gland and are active components of *B. dendrophila* venom. Six unique 3FTx sequences were found in *B. irregularis* venom, but none were 100% identical to either irditoxin subunits (Pawlak et al. 2009), although one sequence was 96% identical to irditoxin subunit B. One of the 3FTx *B. cynodon* clones was 97% identical (amino acid sequence) to irditoxin subunit B and another sequence had 83% amino acid sequence identity with irditoxin subunit A; these toxins also clustered together with irditoxin in the 3FTx tree (Fig. 32). These results suggest that *B. cynodon* venom likely contains a prey-specific heterodimeric 3FTx complex similar to irditoxin.

One unique metalloproteinase sequence was amplified, cloned and sequenced from the venom of the Puerto Rican Racer (*Alsophis portoricensis*). Although this sequence was not similar to alsophinase, a previously characterized metalloproteinase within *A. portoricensis* venom (Weldon and Mackessy 2012), it was similar in sequence to other rear-fanged and Elapidae venomous snake P-III metalloproteinase cDNA sequences (Fig. 33). The complete metalloproteinase sequence was amplified, based on the observed amplified product size, but longer venom protein transcripts (>2,000bp) required multiple sequencing reactions that were not performed for this analysis, and therefore only the partial sequence is shown in the alignment (Fig. 33).

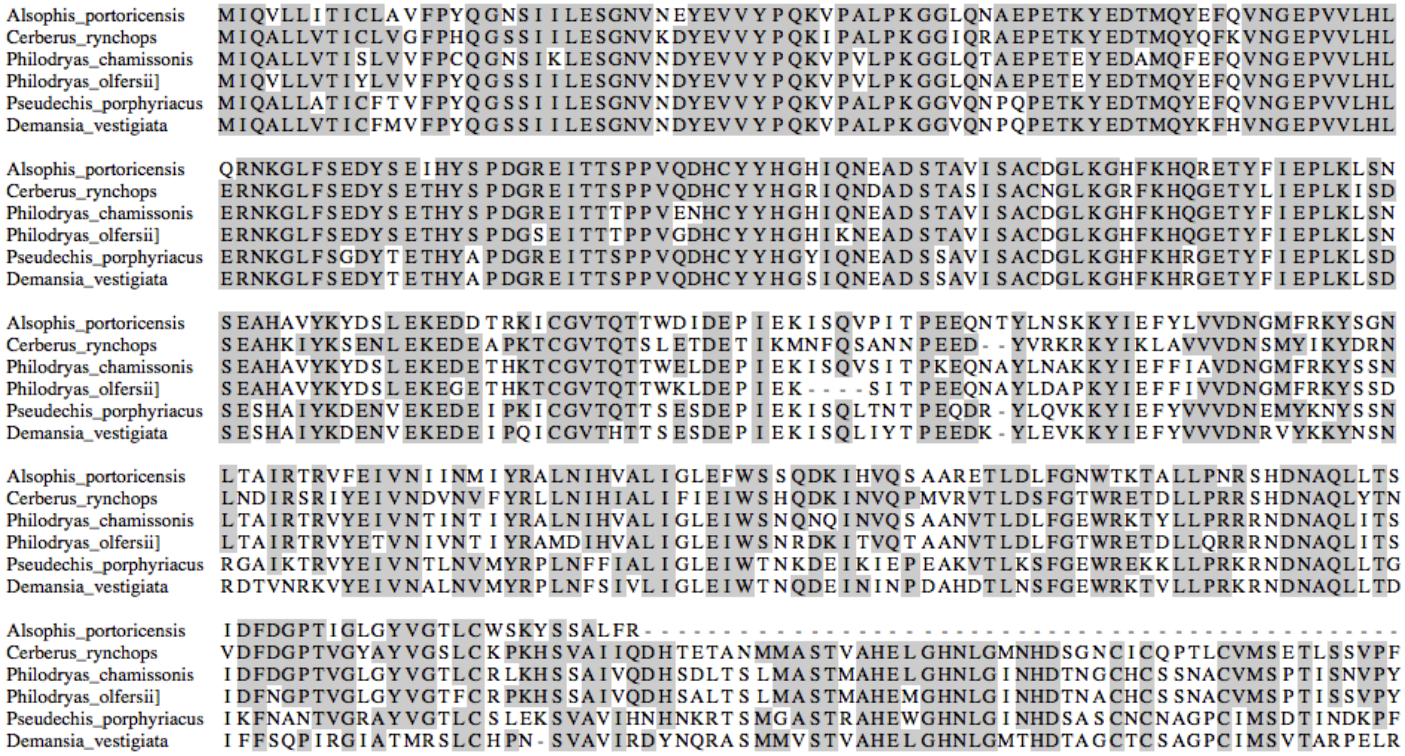


Figure 33. *Alsophis portoricensis* venom PIII metalloproteinase sequence aligned with amino acid sequences from rear-fanged and Elapidae snake species. Identical residues are shaded, demonstrating PIII metalloproteinase sequence conservation in these diverse species. Only a partial sequence of the complete transcript from *A. portoricensis* was used for the alignment. Genbank accession numbers are as follows: Philodryas_chamissonis (AJB84503.1),Philodryas_olfersii (ACS74987.1),Cerberus_rynchops (ADJ51055.1),Pseudechis_porphyriacus (ABQ01133.1), and Demansia_vestigiata (ABK63559.1).

Other cDNA Sequences within Venom

Other sequences incidentally amplified while optimizing primers, from both rattlesnake and rear-fanged snake venom, included complete 60S ribosomal sequences. These sequences, from *C. cerastes* and *A. portoricensis* venom, were 99% identical to the predicted Burmese Python (*Python bivittatus*) 60S ribosomal protein L7a (XP_007420634.1) and L15 isoform X1 (XP_007421748.1), respectively. There were also 40S ribosomal protein sequences amplified from *C. s. scutulatus* which showed 100% sequence identity with the 40S ribosomal protein S9-like isoform X1 from *P. bivittatus* (XP_007439934.1). Cathelicidin-OH antimicrobial peptides (XP_007442672.1) were identified from *C. o. cerberus* and *B. irregularis* venom. These sequences were observed in both rattlesnake and rear-fanged snake venoms, demonstrating that other complete transcripts, in addition to venom protein transcripts, exist within venoms.

Relationship of Sequence Similarities to Empirical and Predictive Protein Activities

Sequences that were similar to crotoxin/Mojave toxin acidic (A) subunits in *C. o. concolor* and *C. simus tzabcan* venoms formed one well-supported clade (1.0 posterior probability), and sequences that were similar to crotoxin/Mojave toxin basic (B) subunits discovered in *C. o. concolor, C. simus tzabcan,* and *C. basiliscus* clustered with other known neurotoxic N6 PLA_2 homologs (1.0 posterior probability) (Fig. 34). Other PLA_2s from *C. o. concolor, C. simus*

tzabcan, and *C. basiliscus* venoms, and also from *C. pricei, C. cerastes, C. m. nigrescens, C. o. cerberus,* and *S. m. barbouri* venoms, clustered within an acidic hemolytic PLA_2 clade shared with other rattlesnakes (0.96 posterior probability) (Fig. 34). It has been experimentally determined that even neurotoxic PLA_2s can also exhibit anticoagulant activity, and this appears to be a common characteristic of many venom PLA_2 enzymes.

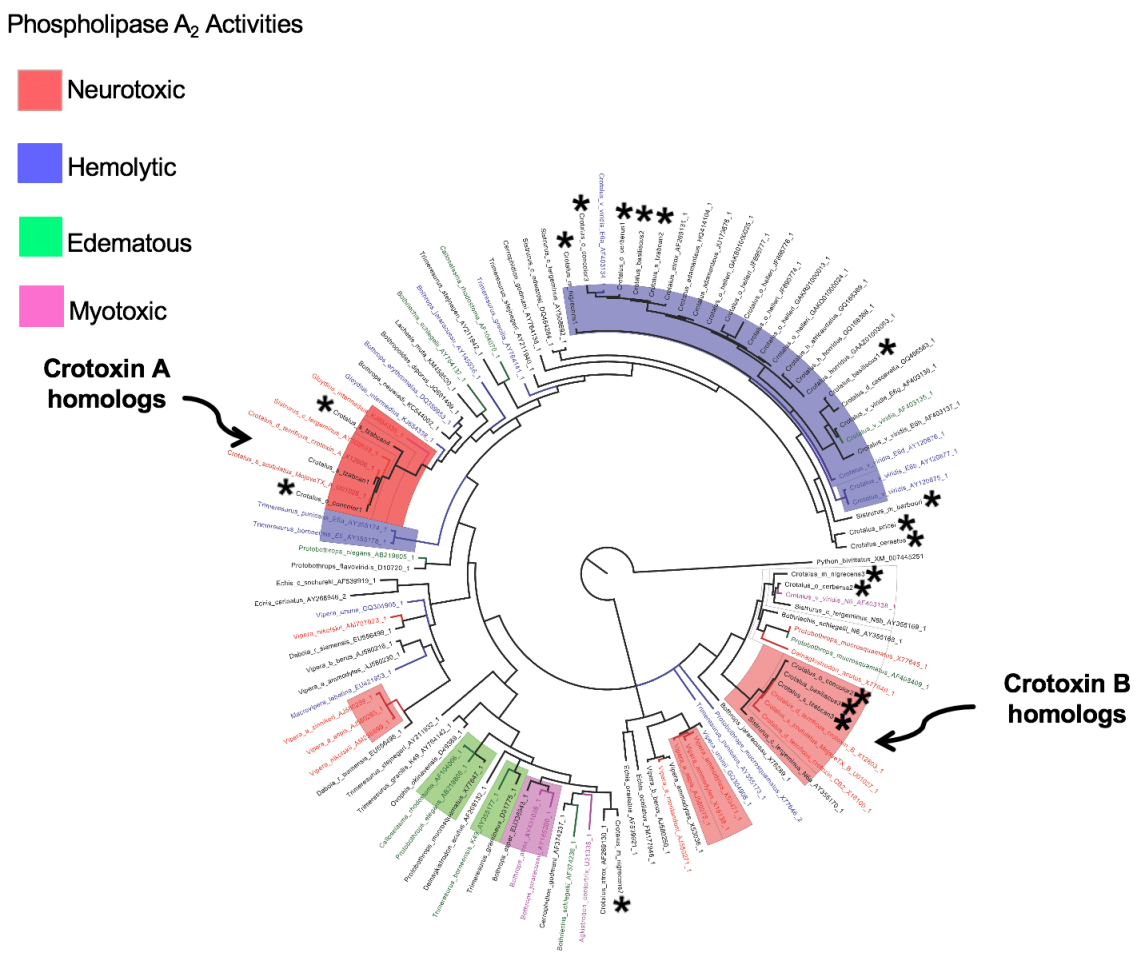


Figure 34. Bayesian sequence similarity tree depicting viper Group II phospholipase A_2 (PLA_2) relationships. PLA_2 cDNA sequences were obtained from *Crotalus simus tzabcan, C. cerastes, C. molossus nigrescens, C. oreganus concolor, C. organus cerberus, C. pricei, C. basiliscus, and Sistrurus miliarius barbouri* venoms (asterisks), and their relationships with various published Old and New World viperid venom PLA_2s are shown. GenBank accession numbers are listed within the tree. Published PLA_2 sequences are colored corresponding to empirically determined biological activity. Whole clades are highlighted where at least two of the published sequences within the clade have the same experimentally determined activity.

Positive Selection Analysis of Rattlesnake Venom PLA_2s and Rear-Fanged Snake Venom 3FTxs

Sites experiencing positive selection were determined by likelihood ratio tests using twice the negative difference in log likelihoods (denoted by Δ), nonsynonymous to synonymous mutation ratios (*dN/dS* = *ω*), and codon-site class frequencies (denoted by *p*) (Table 18). For both venom protein families (rattlesnake PLA_2s and rear-fanged venomous snake 3FTxs), the positive selection model (M2) was found to fit each dataset best in comparison to the nearly neutral model (M1) (Table 18).

Table 18. Summary of codeml tests for positive selection within venom protein superfamilies.

Toxin Family	n/Model	M1: Nearly Neutral	-lnL	M2: Positive Selection	-lnL	M0:	Δ	*p* value
3FTx	34 GTR + I + G	*ρ:* 0.30 0.69 *ω:* 0.14 1.00	5440.65	*ρ:* 0.20 0.31 0.47 *ω:* 0.15 1.00 2.99	5364.39	*ω:* 1.53	152.52	*P* < 0.001*
PLA_2	111 GTR + I + G	*ρ:* 0.37 0.62 *ω:* 0.06 1.00	11332.23	*ρ:* 0.35 0.38 025 *ω:* 0.08 1.00 2.94	11153.15	*ω:* 0.83	358.16	*P* < 0.001*

* indicates significance

Within viper PLA_2s, approximately 20% of the sites were found to be under positive selection, and approximately 28% of sites were functionally critical and conserved (Fig. 35). Models of the basic neurotoxic subunits of Mojave

toxin/crotoxin homologs from *C. o. concolor* and *C. s. tzabcan* (current study) were compared to models of the Mojave toxin basic (B) subunit from *C. s. scutulatus* and the crotoxin basic (B) subunit from *C. d. terrificus* (Fig. 35A-D), in addition to the acidic PLA_2 from *C. pricei* (Fig. 35E) with predicted edematous activity. These comparisons reveal regions of shared protein functional conservation and sensitivity to mutation between all models, regardless of activity.

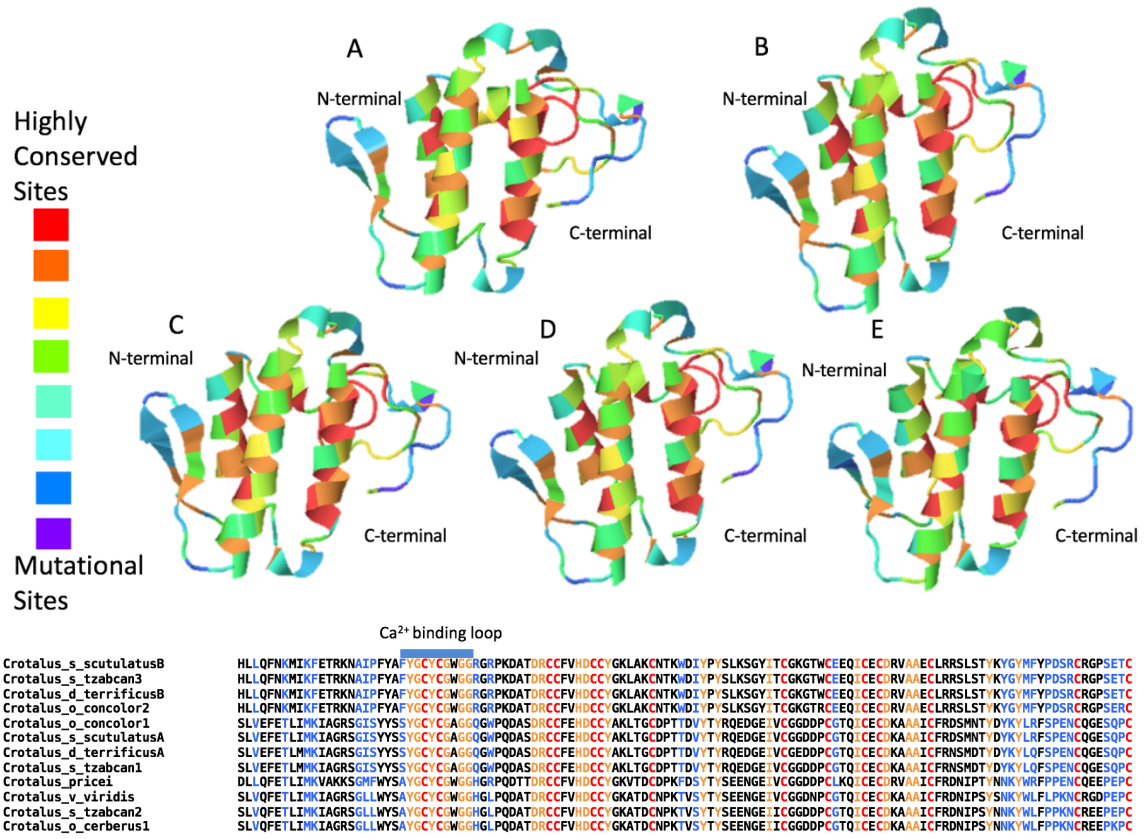


Figure 35. Phospholipase A_2 (PLA_2) structural models and sequence alignments from rattlesnake venoms with conserved and mutational sites. Models of the basic subunits of neurotoxic Mojave toxin/crotoxin homologs of A) *Crotalus oreganus concolor*, B) *Crotalus scutulatus scutulatus*, C) *Crotalus simus tzabcan*, and D) *Crotalus durissus terrificus* are compared to the acidic PLA_2 with predicted edematous activity from E) *Crotalus pricei* to demonstrate overall shared regions of residue conservation and sensitivity to mutation. F) Sequence regions highlighted in red and orange are highly conserved and sensitive to mutations, while regions highlighted in blue are amino acids that are evolving under positive selection and have an increased mutational rate. The conserved calcium binding loop is indicated by the blue bar above the sequence alignment. Genbank accession numbers follow species names for known toxins.

Although both positive selection tests found the positive selection model (M2) to be favored (Table 18), the overall *dN/dS* (*ω*) generated from the M0 model fell under 1 for the PLA_2 superfamily. However, an overall *ω* can be a poor indicator of positive selection because it is averaged across sites. Not all sites are evolving under positive selection; therefore, although the PLA_2 superfamily has sites experiencing positive selection, the number and proportion of functionally conserved and constrained sites appears to be greater than what was observed for the 3FTx superfamily.

Within rear-fanged/non-conventional snake venom 3FTxs, approximately 32% of the sites were under positive selection, and approximately 20% of sites were functionally critical and conserved (Fig. 36). Three-finger toxin models and the predictive program SuSPect (Yates et al. 2014), revealed that functionally conserved protein regions, especially the internal beta-pleated sheets, had the greatest mutational sensitivity, and the exposed residues, especially the extended loops and N and C-termini, were the least sensitive to mutation (Fig. 36A-C). Sites that were the least sensitive to mutation were also found to be under positive selection (Fig. 36D). Models predicted for 3FTxs from *Boiga nigriceps* and *Boiga cynodon* venoms, which showed sequences identity (>80%) to the prey-specific *Boiga irregularis* irditoxin A, also shared the same regions of functional conservation and sensitivity to mutation (Fig. 36), revealing a consistent trend among rear-fanged snake venom 3FTxs.

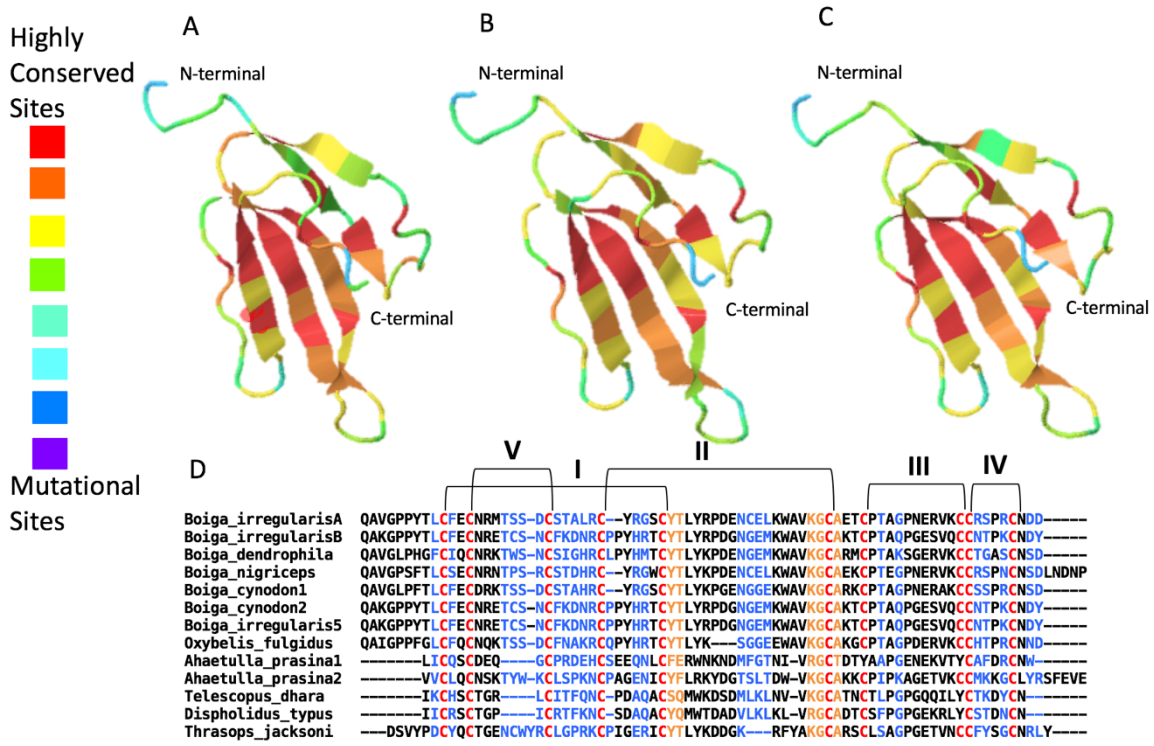


Figure 36. Three-finger toxin (3FTx) structural models and sequence alignments from rear-fanged snake venoms with conserved and mutational sites. Models of 3FTxs with high sequence identity, A) *Boiga nigriceps* and B) *Boiga cynodon1*, are compared with C) prey-specific *Boiga irregularis* irditoxin A to demonstrate regions of shared functional conservation and sensitivity to mutation. These non-conventional 3FTxs all have 10 conserved cysteine residues that form the five disulfide bonds (brackets above sequences) that maintain the “three-finger” structure; N and C-termini are labeled within the models. D) Sequences with residues highlighted in red and orange are highly conserved and sensitive to mutations, while regions highlighted in blue are evolving under positive selection and have an increased mutational rate. Genbank accession numbers of known toxins are as follows: Boiga_irregularisA (DQ304538.1), Boiga_irregularisB (DQ304539.1), Boiga_dendrophila (DQ366293.1), Telescopus_dhara (EU029685.1), Dispholidus_typus (EU029671.1), and Thrasops_jacksoni (EU029674.1).

Discussion and Conclusions

Concentrations of total extracellular RNA within snake venom were observed to be moderate but variable using Nanodrop, and using a Qubit instrument, RNA concentrations were all below instrument detection limit (< 20 ng/µl) (Table 15). Qubit readings provide better accuracy because the fluorescent dye is highly selective for RNA and not DNA, which has also been detected within venom (Pook and McEwing 2005). In addition, common contaminants do not affect Qubit readings. The more accurate Qubit readings revealed a much lower concentration of extracellular RNA within venom, lower than Nanodrop concentrations reported in this and in a previous study (Currier et al. 2012). However, in spite of low RNA concentrations, venom protein cDNAs could still be amplified successfully from both front and rear-fanged venomous snakes. By using extracellular messenger RNA within venom to obtain full-length venom protein transcripts, this method can be used without the need to sacrifice living animals to obtain venom gland tissue. It was also possible to successfully amplify full-length venom protein transcripts from venom that was fresh, lyophilized or stored at -20 °C for 20+ years, desiccated over silica gel in the field, or obtained from a commercial venom supplier. This represents a significant advance over previous attempts to amplify venom-derived mRNAs, which typically produced only partial sequence transcripts (Chen et al. 2002; Currier et al. 2012).

Venom protein genes experience an accelerated rate of nucleotide substitution (Ogawa et al. 1992; Nobuhisa et al. 1996), making it difficult to design sense and antisense pairs of primers to amplify unknown venom proteins

sequences, which is why complete venom gland transcriptomes from gland tissue are usually assembled. However, venom proteins demonstrate high conservation of nucleotide signal peptide sequence and/or 5'UTRs (Figs. 29; 31). By designing degenerate sense primers from these conserved nucleotide sequences and performing 3'RACE, the successful amplification of a diversity of transcript sequences for the major venom protein families (metalloproteinases, serine proteases, C-type lectins, phospholipase A_2s, and three-finger toxins) responsible for clinically significant snakebite were obtained from venom. This approach also allowed determination of unique, currently unknown full-length toxin sequences for many front-fanged and rear-fanged species in all of the major clades of venomous snakes (Viperidae, Elapidae, "Colubridae"). This use of degenerate primers to amplify unknown full-length venom protein sequences within a superfamily from snake venom can be employed to screen sequences within each species for toxins of interest, to examine novel mutations within a venom protein superfamily, or to provide an inexpensive method to obtain complete amino acid sequence for a protein under investigation.

Venom gland transcriptomes generated from next-generation sequencing (Roche 454 or Illumina) provide more comprehensive transcriptome profiles and identify the complete repertoire of transcripts within each venom protein superfamily (Rokyta et al. 2012b; Aird et al. 2013; Rokyta et al. 2013). An abundance of unique 3FTx transcripts identified in rear-fanged snake venom gland transcriptomes generated by next-generation sequencing has been reported, with over fifty 3FTxs transcripts in the case of *B. irregularis* (McGivern

et al. 2014). The number of unique venom protein PLA_2 sequences discovered in viper venom gland transcriptomes completed with next-generation sequencing ranges from 4-9 (Rokyta et al. 2012b; Aird et al. 2013; Rokyta et al. 2013); therefore, the number of unique sequences obtained in this study were by no means a comprehensive evaluation of all transcripts within these protein superfamilies. However, using established procedures which are readily accessible to many researchers, such as 3'RACE and the selection/sequencing of *E. coli* clones, it was possible to identify the major transcripts present for each venom protein superfamily explored in this study. The approach used here allows for researchers interested in a single venom protein superfamily to selectively obtain all highly abundant transcripts for that protein superfamilies, is cost-effective, and does not require the computing resources and bioinformatics needed for next generation sequencing transcriptome assemblies. Because venom protein cDNA sequences are obtained from venoms, this method also allows for the assembly of a genotype-phenotype map, using only venom as source material.

Phospholipase A_2 enzymes and 3FTxs were chosen as the main focus of this study because they constitute very large venom protein superfamilies that exhibit a diversity of activities, including neurotoxic, myotoxic, cardiotoxic, anticoagulant and hemolytic activities (Kini 2003; Kini and Doley 2010). These venom proteins are ideal for structure/function studies as well as protein engineering studies, because a variety of activities and functional sites are possible using the same conserved protein structural scaffold. Also, 3FTxs and

PLA_2s are venom proteins that are observed in abundance in snake venoms, and both are toxins that contribute significantly to serious snake envenomation symptomology. Crotoxin or Mojave toxin PLA_2 heterodimeric complexes result in phenotypically neurotoxic venoms, and the absence or presence of these complexes result in distinctive venom types that have been labeled type I and type II. Type I venoms have higher metalloproteinase activity and lower toxicity, and type II venoms have low metalloproteinase activity and high toxicity/neurotoxicity (Mackessy 2010a).

There can be variation in the occurrence of crotoxin/Mojave toxin complexes within a species, as is seen among different populations of *C. horridus, C. scutulatus* and *C. simus* throughout their range (Glenn et al. 1994; Massey et al. 2012; Castro et al. 2013). This study shows that it is possible to detect the acidic and basic subunit transcripts of these neurotoxic PLA_2 complexes within venom. In the case of the *C. s. tzabcan* utilized in this study, several isoforms of acidic and basic crotoxin-like subunits were observed. The neurotoxicity of *C. s. tzabcan* venom varies with snake locality (Castro et al. 2013); because the specific locality of the *C. s. tzabcan* used in this study was unknown, sequencing venom protein transcripts present within venom was a successful approach to evaluating venom phenotype.

This technique can also be used to analyze the amino acid sequences of toxins in unknown venoms, and this study is the first to report the complete sequence for both subunits of concolor toxin from venom of *C. o. concolor*. Although it has been known for some time that this neurotoxic complex is present

in *C. o. concolor* venom (Glenn and Straight 1977; Weinstein et al. 1985), the presence of acidic and basic crotoxin/Mojave toxin homologs confirmed that a PLA_2-based neurotoxin was present in this type II venom. The viper PLA_2 Bayesian sequence similarity tree revealed some distinctive clusters that corresponded with experimentally characterized PLA_2 protein activities. For example, analysis of PLA_2 sequences from *C. o. cerberus*, a subspecies with type I venom, demonstrated that its PLA_2 clustered within the acidic hemolytic PLA_2 clade, as is typical of many low toxicity rattlesnake PLA_2s. Mojave and crotoxin-like PLA_2 clusters for acidic and basic subunits were also separate from the clades that contained Old World viper neurotoxic PLA_2 complexes (basic and acidic subunits of a heterodimer PLA_2 from *Vipera nikolskii* and vaspin subunits from *Vipera aspis aspis*), suggesting the possibility of a separate evolutionary origin for Old World and New World neurotoxic heterodimeric PLA_2 complexes.

Venom 3FTxs and PLA_2s can have multiple different active sites, and individual toxins are rarely tested for all possible activities or substrates, so it is difficult to predict protein activities or to determine if misclassifications are occurring with predictive methods based solely on sequences similarities (Malhotra et al. 2013). Nevertheless, sequence similarity clustering did successfully identify crotoxin and Mojave toxin homologs, PLA_2s that are associated with serious neurotoxic envenomation symptomology, in known and previously uninvestigated venoms. Two 3FTx transcripts were discovered in *Boiga cynodon* venom that were very similar in sequence to the two subunits of the heterodimeric, prey-specific iriditoxin from *B. irregularis* venom, indicating the

presence of another lizard and bird-specific neurotoxin within the venom of a closely related species. Full-length venom protein transcripts obtained from venom can therefore be used to screen for particular toxins or venom phenotypes.

It is well known that venom proteins experience an accelerated rate of protein evolution, with surface residues undergoing a greater rate of nucleotide substitutions than buried residues (Lynch 2007; Sunagar et al. 2013). Venom genes undergo multiple duplications, which when combined with the higher mutations of exposed residues, results in gene neofunctionalization (Casewell et al. 2013; Vonk et al. 2013). These mechanisms have allowed large venom protein superfamilies (3FTxs, PLA_2s, etc.) to gain a diversity of protein functionalities while maintaining a conserved structural scaffold. The methods detailed here can also be used to evaluate evolutionary pressures shaping venom phenotypes and mechanisms of venom protein adaptive evolution.

Sites in both rattlesnake venom PLA_2s and rear-fanged snake venom 3FTxs that undergo higher mutation rates were identified using positive selection analysis, which identified sites that are sensitive or insensitive to mutation as well as regions conserved by protein structural constraints (Yang 2007; Yates et al. 2014). Both PLA_2s and 3FTxs are known to experience positive selection (Gibbs and Rossiter 2008; Sunagar et al. 2013), but the focus of the present study was to evaluate trends in positively selected and conserved sites across a series of PLA_2s and 3FTxs derived from a diversity of rattlesnake and rear-fanged snake venoms. Analysis of novel rear-fanged snake 3FTxs demonstrated that the same

sites are experiencing positive selection as observed for other 3FTxs previously analyzed (Sunagar et al. 2013; McGivern et al. 2014). For both rear-fanged snake 3FTxs and viper PLA_2s, almost all conserved and mutational sites were shared across each superfamily, regardless of protein activity or abundance in venoms. 3FTxs have a higher percentage of residues that experience positive selection in comparison to PLA_2s, which, though also evolving under positive selection, have a greater percentage of functionally conserved sites. This difference is most likely related to the structural constraints within PLA_2 enzymes that are vital for protein functionality, such as the calcium binding site and the catalytic site responsible for the hydrolytic release of arachidonic acid and lysophospholipids from the sn-2 acyl bond of phospholipids (Kini 2003). There is also considerable secondary structure in the PLA_2 proteins, including multiple α helices and beta-pleated sheets, compared to the few beta-pleated sheets that are observed in 3FTxs. For both protein superfamilies, mutational sites were commonly observed in residues near the ends of loops or the N- and C-termini. Sites that are under positive selection are likely sites involved with the evolution of differential binding affinities. For rear-fanged 3FTxs, it has been suggested that the WAVK motif near the end of the second 3FTx loop is responsible for prey-specific binding (Heyborne and Mackessy 2013), and this region did not appear to be under positive selection following this analysis; however, residues a few amino acids upstream from this site were found to be evolving under positive selection. Characterization of more rear-fanged 3FTxs will be necessary to determine which structural features contribute to prey specificity function.

As more full-length transcripts become available, high throughput methods such as next-generation proteomic (and transcriptomic) characterization of venoms that lack profiles, including most rear-fanged snake species and many understudied front-fanged snake species, will be greatly facilitated. The methods described here provided full-length venom protein transcripts from venoms representing the three major families of venomous snakes, making it is possible to determine snake venom genotype-phenotype relationships without the need to sacrifice living snakes. By requiring only venom to obtain venom protein cDNAs, the approaches detailed here will provide access to cDNA-based protein sequences in the absence of living specimens, from commercial and other venom sources, and will facilitate study of snake venom protein composition and evolution, and in turn, provide greater predictability of the development of regionally-specific reactions following snakebite envenomation.

CHAPTER V

FINAL CONCLUSIONS

Advancements in high-throughput proteomic (tandem mass spectrometry) and genomic (DNA sequencing) technologies have resulted in the ability to generate comprehensive venom profiles for many front and rear-fanged venomous snake species. Rear-fanged snake venom research has slowly progressed due to the difficulties obtaining crude venom and a lack of interest in snakes that only rarely are responsible for human morbidity and mortality (Mackessy 2002; Modahl et al. 2015). However, current research into rear-fanged snake venoms has demonstrated the existence of novel venom proteins and has provided insight into the evolution and origin of snake venom toxins within advanced snakes (OmPraba et al. 2010; Ching et al. 2012). These venoms still remain largely unexplored, and there exists within these venoms the potential to discover proteins of therapeutic significance or with unique characteristics, an example being the evolution of prey-specific toxicity (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013).

Rear-fanged snake venoms are currently the only known venoms with prey-specific toxins. Venom proteins that have been observed to exhibit prey-specific toxicity belong to the three-finger toxin (3FTx) family (Pawlak et al. 2006; Pawlak et al. 2009; Heyborne and Mackessy 2013). Interestingly, although some

3FTxs are prey-specific, others are non-toxic or are toxic to a diversity of prey types (not prey-specific) (Nirthanan et al. 2003). *Naja kaouthia* has venom that is dominated by many 3FTx isoforms, but only one 3FTx is lethal (<1 μg/g) toward both ectothermic (lizard) and endothermic (mammal) prey. As a generalist predator, the *N. kaouthia* consumes a variety of prey, and a "generalist" 3FTx that incapacitates a large diversity of prey types is the main lethal toxin that has evolved in this venom.

Pseustes sulphureus, which also has a venom dominated by multiple 3FTx isoforms, has evolved a different trophic adaptation with two prey-specific 3FTxs within its venom arsenal. The most abundant toxin (sulmotoxin A), a heterodimeric 3FTx complex, demonstrates lizard-specific toxicity. The second most abundant 3FTx (sulmotoxin B) is non-toxic towards lizards, but is toxic (<5 μg/g) towards mice and contributes to the overall toxicity of crude *P. sulphureus* venom toward mammals. *Pseustes sulphureus* was also found to produce one of the more toxic rear-fanged snake venoms (Mackessy 2002; Mackessy et al. 2006; Weldon and Mackessy 2010; Weinstein et al. 2011). Having two prey-specific 3FTxs within this venom allows for the incapacitation of a diversity of prey types, both ectothermic and endothermic. The venoms of the *N. kaouthia* and *P. sulphureus* demonstrate how a crude venom dominated by one protein superfamily (3FTxs) can still allow for the targeting of a diversity of prey.

Three-finger toxins are ideal to study protein structure and function relationships. Besides prey-specific neurotoxicity, 3FTxs can also display cardiotoxic, hemorrhagic, or tissue necrotizing effects. However, all 3FTxs share

a conserved structural scaffold with three β-stranded loops extending from a small, globular, hydrophobic core that is crossed linked by four conserved disulfide bridges (Kini and Doley 2010). Current published lizard-specific 3FTxs were discovered to share a conserved amino acid sequence, WAVK, within the second 3FTx loop, and this sequence of amino acids has been suggested to be associated with receptor binding and confer taxon-specific neurotoxicity (Heyborne and Mackessy 2013). Further, sulmotoxin A, the lizard-specific heterodimeric 3FTx complex, also shared this amino acid sequence. Sulmotoxin B, which was mammal specific in toxicity, exhibited an alanine to threonine substitution within this region. This dissertation work provides a foundation to begin to explore the amino acids responsible for 3FTx receptor-binding specificity. These toxins are highly stable scaffolds with potential use in therapeutic protein engineering (e.g., painkiller development) if their receptor binding affinities can be modified and directed to specific targets (Gilquin et al. 2003; Nirthanan and Gwee 2004; Lewis 2009).

Three-finger toxins have already provided insight into the distribution of specific receptors or ion channels in particular tissues or cells, as well as in imaging of receptor trafficking (Tsetlin 1999; Nirthanan and Gwee 2004; Kini and Doley 2010). For example, α-bungarotoxin, a canonical 3FTx, has played a fundamental role in the isolation and characterization of the nicotinic acetylcholine receptor at the motor end plate (Yee et al. 2004). Because of the specific binding affinities of α-bungarotoxin, the nicotinic acetylcholine receptor is one of the most thoroughly characterized neuronal receptors today (Nirthanan

and Gwee 2004). Therefore, exploring the activities of non-conventional 3FTxs in elapid and rear-fanged snake venoms provides insight into therapeutic protein engineering and/or receptor characterization.

Research documenting overall snake venom composition is important for identifying venoms with protein targets for therapeutic engineering, as well as identifying snakes that could be potentially hazardous to human health if bitten. As previously mentioned, many rear-fanged snakes have currently unexplored venoms and therefore, the risk they pose to humans is unknown. The most comprehensive method to characterize a snake venom to date includes a proteomics approach with MS/MS (tandem mass spectrometry) peptide sequencing of separated venom components, combined with a species-specific venom gland transcriptome (Aird et al. 2013; Sunagar et al. 2014; Aird et al. 2015).

Using MS/MS identification of peptide sequences and relying on online protein sequence databases alone can overlook unique isoform variations and can fail to recognize novel venom proteins, because only small peptide fragments of a portion of the molecule are used for protein identification. By generating a complementary transcriptome for a snake species, MS/MS peptide sequences can be more precisely identified to their corresponding transcript. There is a lack of venom proteins and transcripts from rear-fanged venomous snakes within current databases, and this has resulted in the inability to identified prominent venom proteins within these venoms using only available online protein resources.

This dissertation demonstrated the critical importance of using a species-specific transcriptome in combination with MS/MS to document the venom composition of rear-fanged venomous snakes. The combination of a species-specific transcriptome with MS/MS provided complete overall venom composition (all proteins present were able to be identified) and was also useful to determine the relative abundances of individual venom proteins and venom protein superfamilies. Further, this approach could be useful for other fields, especially research involving non-model organisms, where *de novo* transcriptomes and proteomes are assembled.

Venom gland transcriptomes that have been assembled to obtain toxin transcripts have in the past required the sacrifice of live snakes for venom gland tissue. Extracellular toxin transcripts have been documented within crude venom (Chen et al. 2002; Currier et al. 2012), but these studies did not target novel sequences (all sequences were previously known) and only amplified small regions of transcripts. This dissertation demonstrated that full-length venom protein transcripts are present within crude snake venoms (for all major venomous snake families) and that with the use of degenerate primer design, novel venom protein transcripts within each venom protein superfamily can be obtained. This provides an alternative method to acquire venom protein transcripts and information about full-length venom protein amino acid sequences. Further, this approach provides the opportunity to generate both venom transcriptomic and proteomic profiles, essentially a genotype-phenotype

map, using only the same venom sample from one individual. This unique approach has the potential to advance current venom research greatly.

Whole venom characterization should not only be based on the proteins and/or transcripts that are present within a venom or a venom gland, but also on the biological roles these proteins provide within a venom. With the addition of assays to detect venom protein activities and functionalities, predictions regarding venom protein biological roles can be developed and venoms can be more extensively characterized. This dissertation provided examples of an integrated approach to study rear-fanged snake venoms, from the transcript and protein sequences to protein activities and their relationship to snake diet (trophic adaptation). Evaluating the biological roles that venom proteins provide, such as prey-specific targeting, in combination with a transcriptome and proteome, allows for the identification of amino acids responsible for specific receptor binding and provides insight into the abundance, diversity, and evolution of these proteins. It is critical to study the venoms of these most diverse clades of venomous snakes, the rear-fanged “colubrid” snakes, to understand the multiple factors driving the evolution of venoms in squamate reptiles. This dissertation represents an important step towards that goal with the incorporation of multiple high-throughput characterization methods and the addition of biological assays to explore the evolution and biological roles 3FTxs provide within venoms.

REFERENCES

Acosta, O., L. C. Leiva, M. E. Peichoto, S. Marunak, P. Teibler, and L. Rey. 2003. Hemorrhagic activity of the Duvernoy's gland secretion of the xenodontine colubrid *Philodryas patagoniensis* from the north-east region of Argentina. Toxicon 41:1007-1012.

Aird, S. D., S. Aggarwal, A. Villar-Briones, M. M.-Y. Tin, K. Terada, and A. S. Mikheyev. 2015. Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. BMC Genomics 16:647.

Aird, S. D. and N. J. da Silva. 1991. Comparative enzymatic composition of Brazilian coral snake (*Micrurus*) venoms. Comparative Biochemistry and Physiology 99B:287-294.

Aird, S. D., W. G. Kruggel, and I. I. Kaiser. 1990. Amino acid sequence of the basic subunit of Mojave toxin from the venom of the Mojave rattlesnake (*Crotalus s. scutulatus*). Toxicon 28:669-673.

Aird, S. D., Y. Watanabe, A. Villar-Briones, M. C. Roy, K. Terada, and A. S. Mikheyev. 2013. Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (*Ovophis okinavensis* and *Protobothrops flavoviridis*). BMC Genomics 14.

Alape-Giron, A., L. Sanz, J. Escolano, M. Flores-Diaz, M. Madrigal, M. Sasa, and J. J. Calvete. 2008. Snake venomics of the lancehead pitviper *Bothrops asper* geographic, individual, and ontogenetic variations. Journal of Proteome Research 7:3556-3571.

Andrade, D. V. and A. S. Abe. 1999. Relationship of venom ontogeny and diet in *Bothrops*. Herpetologica 55:200-204.

Asad, M. H., M. T. Razi, T. Khan, Q. Najam-Us-saqib, G. Murtaza, M. S. Hussain, S. Karim, and I. Hussain. 2012. Coagulopathies in *Naja naja karachiensis* (black Pakistan cobra) bites and its effect on coagulation tests upon storage of platelet-poor plasma. Acta Poloniae Pharmaceutica 69:1031-1034.

Bandeira, N., K. R. Clauser, and P. A. Pevzner. 2007. Shotgun protein sequencing: Assembly of peptide tandem mass spectra from mixtures of modified proteins. Molecular & Cellular Proteomics 6:1123-1134.

Bieber, A. L., R. R. Becker, R. McParland, D. F. Hunt, J. Shabanowitz, J. R. Yates Iii, P. A. Martino, and G. R. Johnson. 1990. The complete sequence of the acidic subunit from Mojave toxin determined by Edman degradation and mass spectrometry. Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology 1037:413-421.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72:248-254.

Calvete, J. J. 2013. Snake venomics: From the inventory of toxins to biology. Toxicon 75:44-62.

Calvete, J. J. 2014. Next-generation snake venomics: Protein-locus resolution through venom proteome decomplexation. Expert Review of Proteomics 11:315-329.

Calvete, J. J., P. Ghezellou, O. Paiva, T. Matainaho, A. Ghassempour, H. Goudarzi, F. Kraus, L. Sanz, and D. J. Williams. 2012. Snake venomics of two poorly known Hydrophiinae: Comparative proteomics of the venoms of terrestrial *Toxicocalamus longissimus* and marine *Hydrophis cyanocinctus*. Journal of Proteomics 75:4091-4101.

Calvete, J. J., L. Sanz, Y. Angulo, B. Lomonte, and J. M. Gutierrez. 2009. Venoms, venomics, antivenomics. FEBS Letters 583:1736-1743.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. BMC bioinformatics 10:421-421.

Capra, J. A. and M. Singh. 2007. Predicting functionally important residues from sequence conservation. Bioinformatics (Oxford, England) 23:1875-1882.

Cardoso, K. C., M. J. Da Silva, G. G. L. Costa, T. T. Torres, L. E. V. Del Bem, R. O. Vidal, M. Menossi, and S. Hyslop. 2010. A transcriptomic analysis of gene expression in the venom gland of the snake *Bothrops alternatus* (urutu). BMC Genomics 11.

Casewell, N. R., G. A. Huttley, and W. Wuster. 2012. Dynamic evolution of venom proteins in squamate reptiles. Nature communications 3:1066.

Casewell, N. R., S. C. Wagstaff, R. A. Harrison, C. Renjifo, and W. Wüster. 2011. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Molecular Biology and Evolution 28:2637-2649.

Casewell, N. R., S. C. Wagstaff, W. Wüster, D. A. N. Cook, F. M. S. Bolton, S. I. King, D. Pla, L. Sanz, J. J. Calvete, and R. A. Harrison. 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proceedings of the National Academy of Sciences 111:9205-9210.

Casewell, N. R., W. Wuster, F. J. Vonk, R. A. Harrison, and B. G. Fry. 2013. Complex cocktails: The evolutionary novelty of venoms. Trends Ecology and Evolution 28:219-229.

Castoe, T. A., A. P. J. de Koning, K. T. Hall, D. C. Card, D. R. Schield, M. K. Fujita, R. P. Ruggiero, J. F. Degner, J. M. Daza, G. Wanjun, J. Reyes-Velasco, K. J. Shaney, J. M. Castoe, S. E. Fox, A. W. Poole, D. Polanco, J. Dobry, M. W. Vandewege, L. Qing, and R. K. Schott. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proceedings of the National Academy of Sciences of the United States of America 110:20645-20650.

Castro, E. N., B. Lomonte, M. Del Carmen Gutiérrez, A. Alagón, and J. M. Gutiérrez. 2013. Intraspecies variation in the venom of the rattlesnake *Crotalus simus* from Mexico: Different expression of crotoxin results in highly variable toxicity in the venoms of three subspecies. Journal of Proteomics 87:103-121.

Chaitae, A. 2000. Demography of the monocled cobra (*Naja kaouthia*) in the central region of Thailand. Department of Biology. University of Louisville, Louisville, Kentucky.

Chanda, C., A. Sarkar, and D. Chakrabarty. 2015. Thrombolytic protein from cobra venom with anti-adhesive properties. Archives of Biochemistry and Biophysics 590:20-26.

Chang, H. C., T. S. Tsai, and I. H. Tsai. 2013. Functional proteomic approach to discover geographic variations of king cobra venoms from Southeast Asia and China. Journal of Proteomics 89:141-153.

Chapeaurouge, A., M. A. Reza, S. P. Mackessy, P. C. Carvalho, R. H. Valente, A. Teixeira-Ferreira, J. Perales, Q. Lin, and R. M. Kini. 2015. Interrogating the venom of the viperid snake *Sistrurus catenatus edwardsii* by a combined approach of electrospray and MALDI mass spectrometry. PLOS One 10.

Chatrath, S. T., A. Chapeaurouge, Q. S. Lin, T. K. Lim, N. Dunstan, P. Mirtschin, P. P. Kumar, and R. M. Kini. 2011. Identification of novel proteins from the venom of a cryptic snake *Drysdalia coronoides* by a combined transcriptomics and proteomics approach. Journal of Proteome Research 10:739-750.

Chen, T. B., A. J. Bjourson, D. F. Orr, H. Kwok, P. F. Rao, C. Ivanyi, and C. Shaw. 2002. Unmasking venom gland transcriptomes in reptile venoms. Analytical Biochemistry 311:152-156.

Chen, Y.-H., Y.-M. Wang, M.-J. Hseu, and I.-H. Tsai. 2004. Molecular evolution and structure–function relationships of crotoxin-like and asparagine-6-containing phospholipases A_2 in pit viper venoms. Biochemical Journal 381:25-34.

Ching, A. T., A. F. Paes Leme, A. Zelanis, M. M. Rocha, F. Furtado Mde, D. A. Silva, M. R. Trugilho, S. L. da Rocha, J. Perales, P. L. Ho, S. M. Serrano, and I. L. Junqueira-de-Azevedo. 2012. Venomics profiling of *Thamnodynastes strigatus* unveils matrix metalloproteinases and other novel proteins recruited to the toxin arsenal of rear-fanged snakes. Journal of Proteome Research 11:1152-1162.

Ching, A. T. C., M. M. T. Rocha, A. F. Paes Leme, D. C. Pimenta, M. de Fátima D. Furtado, S. M. T. Serrano, P. L. Ho, and I. L. M. Junqueira-de-Azevedo. 2006. Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome. FEBS Letters 580:4417-4422.

Chippaux, J. P., V. Williams, and J. White. 1991. Snake venom variability: Methods of study, results, and interpretation. Toxicon 29:1279-1303.

Chiszar, D. and H. M. Smith. 2002. Colubrid envenomations in the United States. Journal of Toxicology-Toxin Reviews 21:85-104.

Currier, R., C. Juan, S. Libia, H. Robert , R. Paul, and W. Simon. 2012. Unusual stability of messenger RNA in snake venom reveals gene expression dynamics of venom replenishment. PLOS One 7:1-10.

da Rocha, M. M. T. and M. D. D. Furtado. 2007. Analysis of biological activities from *Philodryas olfersii* (Lichtenstein) and *P. patagoniensis* (Girard) venoms (Serpents, Colubridae). Revista Brasileira De Zoologia 24:410-418.

da Silva, N. J. and S. D. Aird. 2001. Prey specificity, comparative lethality and compositional differences of coral snake venoms. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 128:425-456.

Daltry, J. C., G. Ponnudurai, C. K. Shin, N. H. Tan, R. S. Thorpe, and W. Wolfgang. 1996a. Electrophoretic profiles and biological activities: Intraspecific variation in the venom of the Malayan pit viper (*Calloselasma rhodostoma*). Toxicon 34:67-79.

Daltry, J. C., W. Wüster, and R. S. Thorpe. 1996b. Diet and snake venom evolution. Nature 379:537-540.

Das, D., N. Urs, V. Hiremath, B. S. Vishwanath, and R. Doley. 2013. Biochemical and biological characterization of *Naja kaouthia* venom from North-East India and its neutralization by polyvalent antivenom. Journal of Venom Research 4:31-38.

Debnath, A., A. Saha, A. Gomes, S. Biswas, P. Chakrabarti, B. Giri, A. K. Biswas, S. D. Gupta, and A. Gomes. 2010. A lethal cardiotoxic–cytotoxic protein from the Indian monocellate cobra (*Naja kaouthia*) venom. Toxicon 56:569-579.

Doley, R., G. F. King, and A. K. Mukherjee. 2004. Differential hydrolysis of erythrocyte and mitochondrial membrane phospholipids by two phospholipase A_2 isoenzymes (NK-PLA_2-I and NK-PLA_2-II) from the venom of the Indian monocled cobra *Naja kaouthia*. Archives of Biochemistry and Biophysics 425:1-13.

Doley, R., S. P. Mackessy, and R. M. Kini. 2009. Role of accelerated segment switch in exons to alter targeting (ASSET) in the molecular evolution of snake venom proteins. BMC Evolutionary Biology 9.

Doley, R. and A. K. Mukherjee. 2003. Purification and characterization of an anticoagulant phospholipase A_2 from Indian monocled cobra (*Naja kaouthia*) venom. Toxicon 41:81-91.

Doley, R., S. Pahari, S. P. Mackessy, and R. M. Kini. 2008. Accelerated exchange of exon segments in Viperid three-finger toxin genes (*Sistrurus catenatus edwardsii*; Desert Massasauga). BMC Evolutionary Biology 8.

Durban, J., P. Juárez, Y. Angulo, B. Lomonte, M. Flores-Diaz, A. Alape-Girón, M. Sasa, L. Sanz, J. M. Gutiérrez, J. Dopazo, A. Conesa, and J. J. Calvete. 2011. Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. BMC Genomics 12:259-259.

Durban, J., A. Perez, L. Sanz, A. Gomez, F. Bonilla, S. Rodriguez, D. Chacon, M. Sasa, Y. Angulo, J. M. Gutierrez, and J. J. Calvete. 2013. Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. BMC Genomics 14.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Ellman, G. L., D. K. Courtney, j. V. Andres, and R. M. Featherstone. 1961. A new and rapid colorimetric determination of acetylcholinesterase activity. Biochemical Pharmacology 7:88-95.

Estrella, A., E. E. Sanchez, J. A. Galan, W. A. Tao, B. Guerrero, L. F. Navarrete, and A. Rodriguez-Acosta. 2011. Characterization of toxins from the broad-banded water snake *Helicops angulatus* (Linnaeus, 1758): isolation of a cysteine-rich secretory protein, helicopsin. Arch. Toxicol. 85:305-313.

Fagundes, F. H. R., M. Oliveira, S. Huancahuire-Vega, F. F. Romero-Vargas, L. A. Ponce-Soto, and S. Marangoni. 2010. cDNA and deduced primary structure of basic phospholipase A_2 with neurotoxic activity from the venom secretion of the *Crotalus durissus collilineatus* rattlesnake. Brazilian Journal of Medical and Biological Research 43:262-270.

Finney, D. J. 1978. Statistical method in biological assay. Charles Griffin & Co., London and High Wycombe.

Fox, J. W. and S. M. Serrano. 2010. Snake venom metalloproteinases *in* S. P. Mackessy, ed. Handbook of Venoms and Toxins of Reptiles. CRC Press/Taylor & Francis Group, Boca Raton, FL.

Francischetti, I. M. B., V. My-Pham, J. Harrison, M. K. Garfield, and J. M. C. Ribeiro. 2004. *Bitis gabonica* (Gaboon viper) snake venom gland: Toward a catalog for the full-length transcripts (cDNA) and proteins. Gene 337:55-69.

Fry, B. G. 2005. From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Research 15:403-420.

Fry, B. G., N. G. Lumsden, W. Wuster, J. C. Wickramaratna, W. C. Hodgson, and R. M. Kini. 2003a. Isolation of a neurotoxin (alpha colubritoxin) from a nonvenomous colubrid: Evidence for early origin of venom in snakes. Journal of Molecular Evolution 57:446-452.

Fry, B. G., H. Scheib, I. d. L. M. Junqueira de Azevedo, D. A. Silva, and N. R. Casewell. 2012. Novel transcripts in the maxillary venom glands of advanced snakes. Toxicon 59:696-708.

Fry, B. G., W. Wuster, R. M. Kini, V. Brusic, A. Khan, D. Venkataraman, and A. P. Rooney. 2003b. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. Journal of Molecular Evolution 57:110-129.

Fry, B. G., W. Wüster, S. F. Ryan Ramjan, T. Jackson, P. Martelli, and R. M. Kini. 2003c. Analysis of Colubroidea snake venoms by liquid chromatography with mass spectrometry: Evolutionary and toxinological implications. Rapid Communications in Mass Spectrometry 17:2047-2062.

Fryklund, L. and D. Eaker. 1975. The complete amino acid sequence of a cardiotoxin from the venom of *Naja naja* (Cambodian Cobra). Biochemistry 14:2860-2865.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England) 28:3150-3152.

García-Gubern, C., R. Bello, V. Rivera, A. Rocafort, L. Colon-Rolon, and H. Acosta-Tapia. 2010. Is the Puerto Rican Racer, *Alsophis portoricensis*, really harmless? A case report series. Wilderness and Environmental Medicine 21:353-356.

Gibbs, H. L. and W. Rossiter. 2008. Rapid evolution by positive selection and gene gain and loss: PLA_2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. Journal of Molecular Evolution 66:151-166.

Gilquin, B., M. Bourgoin, R. Menez, M. H. Le Du, D. Servent, S. Zinn-Justin, and A. Menez. 2003. Motions and structural variability within toxins: Implication for their use as scaffolds for protein engineering. Protein Science 12:266-277.

Glenn, J. L. and R. Straight. 1977. The midget faded rattlesnake (*Crotalus viridis concolor)* venom: Lethal toxicity and individual variability. Toxicon 15:129-132.

Glenn, J. L., R. Straight, and Wolt. 1994. Regional variation in the presence of canebrake toxin in *Crotalus horridus* venom. Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology 107:337-346.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29:644-652.

Greene, H. W. 1983. Dietary correlates of the origin and radiation of snakes. American Zoologist 23:431-441.

Gutiérrez, J. M., M. C. dos Santos, M. de Fatima Furtado, and G. Rojas. 1991. Biochemical and pharmacological similarities between the venoms of newborn *Crotalus durissus durissus* and adult *C. d. terrificus* rattlesnakes. Toxicon 29:1273-1277.

Gutiérrez, J. M., D. Williams, H. W. Fan, and D. A. Warrell. 2010. Snakebite envenoming from a global perspective: Towards an integrated approach. Toxicon 56:1223-1235.

Hamako, J., T. Matsui, S. Nishida, S. Nomura, Y. Fujimura, M. Ito, Y. Ozeki, and K. Titani. 1998. Purification and characterization of kaouthiagin, a von Willebrand factor-binding and -cleaving metalloproteinase from *Naja kaouthia* cobra venom. Thrombosis and Haemostasis 80:499-505.

Han, X., A. Aslanian, and J. R. Yates. 2008. Mass spectrometry for proteomics. Current Opinion in Chemical Biology 12:483-490.

Hargreaves, A. D. and J. F. Mulley. 2015. Assessing the utility of the Oxford Nanopore MinION for snake venom gland cDNA sequencing. PeerJ 3:e1441.

Hargreaves, A. D., M. T. Swain, M. J. Hegarty, D. W. Logan, and J. F. Mulley. 2014. Restriction and recruitment – gene duplication and the origin and evolution of snake venom toxins. Genome Biology and Evolution 8:2088-2095.

Harrison, R. A., D. Cook, C. Renjifo, N. R. Casewell, R. Currier, and S. C. Wagstaff. 2011. Research strategies to improve snakebite treatment: Challenges and progress. Journal of Proteomics 74:1768-1780.

Harrison, R. A., F. Ibison, D. Wilbraham, and S. C. Wagstaff. 2007. Identification of cDNAs encoding viper venom hyaluronidases: Cross-generic sequence conservation of full-length and unusually short variant transcripts. Gene 392:22-33.

He, Y., J. F. Gao, L. H. Lin, X. M. Ma, and X. Ji. 2014. Age-related variation in snake venom: Evidence from two snakes (*Naja atra* and *Deinagkistrodon acutus*) in Southeastern China. Asian Herpetological Research 5:119-127.

Heyborne, W. H. and S. P. Mackessy. 2013. Identification and characterization of a taxon-specific three-finger toxin from the venom of the Green Vinesnake (*Oxybelis fulgidus*; family Colubridae). Biochimie 95:1923-1932.

Hill, R. E. and S. P. Mackessy. 1997. Venom yields from several species of colubrid snakes and differential effects of ketamine. Toxicon 35:671-678.

Hill, R. E. and S. P. Mackessy. 2000. Characterization of venom (Duvernoy's secretion) from twelve species of colubrid snakes and partial sequence of four venom proteins. Toxicon 38:1663-1687.

Huang, P. and S. P. Mackessy. 2004. Biochemical characterization of phospholipase A_2 (trimorphin) from the venom of the Sonoran Lyre Snake *Trimorphodon biscutatus lambda* (family Colubridae). Toxicon 44:27-36.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford, England) 17:754-755.

Jackson, K. 2003. The evolution of venom-delivery systems in snakes. Zoological Journal of the Linnean Society 137:337-354.

Joubert, F. J. and N. Taljaard. 1980a. The complete primary structures of three cytotoxins (CM-6, CM-7 and CM-7A) from *Naja naja kaouthia* (Siamese cobra) snake venom. Toxicon 18:455-467.

Joubert, F. J. and N. Taljaard. 1980b. Purification, some properties and amino-acid sequences of two phospholipases A (CM-II and CM-III) from *Naja naja kaouthia* venom. FEBS Journal 112:493-499.

Joubert, F. J. and N. Taljaard. 1980c. Snake venoms amino acid sequences of 2 melanoleuca-type toxins Hoppe-Seylers Zeitschrift Fur Physiologische Chemie 361:425-436.

Junqueira-de-Azevedo, I. L. M., A. T. C. Ching, E. Carvalho, F. Faria, M. Y. Nishiyama, P. L. Ho, and M. R. V. Diniz. 2006. *Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of cobra (Elapidae) venoms: Implications for snake toxin repertoire evolution. Genetics 173:877-889.

Kamiguti, A. S., R. D. Theakston, N. Sherman, and J. W. Fox. 2000. Mass spectrophotometric evidence for P-III/P-IV metalloproteinases in the venom of the Boomslang (*Dispholidus typus*). Toxicon 38:1613-1620.

Kardong, K. V. 1996. Snake toxins and venoms: An evolutionary perspective. Herpetologica 52:36-46.

Kardong, K. V. 2002. Colubrid snakes and Duvernoy's "venom" glands. Journal of Toxicology-Toxin Reviews 21:1-19.

Kardong, K. V., T. L. Kiene, and V. Bels. 1997. Evolution of trophic systems in squamates. Netherlands Journal of Zoology 47:411-427.

Kardong, K. V. and P. A. Lavin-Murcio. 1993. Venom delivery of snakes as high-pressure and low-pressure systems. Copeia 1993:644-650.

Karlsson, E. 1973. Chemistry of some potent animal toxins. Experientia 29:1319-1327.

Karlsson, E. and D. Eaker. 1972. Isolation of the principal neurotoxins of *Naja naja* subspecies from the Asian mainland. Toxicon 10:217-225.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols 10:845-858.

Kini, R. M. 2003. Excitement ahead: structure, function and mechanism of snake venom phospholipase A_2 enzymes. Toxicon 42:827-840.

Kini, R. M. and R. Doley. 2010. Structure, function and evolution of three-finger toxins: Mini proteins with multiple targets. Toxicon 56:855-867.

Kuch, U. and D. Mebs. 2002. Envenomations by colubrid snakes in Africa, Europe, and the Middle East. Journal of Toxicology-Toxin Reviews 21:159-179.

Kukhtina, V. V., C. Weise, A. V. Osipov, V. G. Starkov, M. I. Titov, S. E. Esipov, T. V. Ovchinnikova, V. I. Tsetlin, and Y. N. Utkin. 2000. The MALDI mass spectrometry in the identification of new proteins in snake venoms. Bioorganicheskaya Khimiya 26:803-807.

Kularatne;, S. A. M., B. D. S. S. Budagoda;, I. B. Gawarammana;, and W. K. S. Kularatne;. 2009. Epidemiology, clinical profile and management issues of cobra (*Naja naja*) bites in Sri Lanka: first authenticated case series. Transactions of the Royal Society of Tropical Medicine and Hygiene 103:924-930.

Kulkeaw, K., W. Chaicumpa, Y. Sakolvaree, P. Tongtawe, and P. Tapchaisri. 2007. Proteome and immunome of the venom of the Thai cobra, *Naja kaouthia*. Toxicon 49:1026-1041.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29:1695-1701.

Laskowski, M. S. 1980. Purification and properties of venom phosphodiesterase. Methods in enzymology 65:276-284.

Laustsen, A. H., J. M. Gutiérrez, B. Lohse, A. R. Rasmussen, J. Fernández, C. Milbo, and B. Lomonte. 2015. Snake venomics of monocled cobra (*Naja kaouthia*) and investigation of human IgG response against venom toxins. Toxicon 99:23-35.

Lewis, R. J. 2009. Conotoxin venom peptide therapeutics. Pp. 44-48. Pharmaceutical Biotechnology.

Li, B. and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12:323.

Li, M., B. G. Fry, and R. M. Kini. 2005a. Eggs-only diet: Its implications for the toxin profile changes and ecology of the marbled sea snake (*Aipysurus eydouxii*). Journal of Molecular Evolution 60:81-89.

Li, M., B. G. Fry, and R. M. Kini. 2005b. Putting the brakes on snake venom evolution: The unique molecular evolutionary patterns of *Aipysuras eydouxii* (Marbled sea snake) phospholipase A_2 toxins. Molecular Biology and Evolution 22:934-941.

Li, W. and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics (Oxford, England) 22:1658-1659.

Lim, K. P. and L. K. Lim. 1992. A guide to the amphibians & reptiles of Singapore. Singapore Science Centre.

Lopez, T. J. and L. R. Maxson. 1995. Mitochondrial DNA sequence variation and genetic differentiation among colubrine snakes (Reptilia: Colubridae: Colubrinae). Biochemical Systematics and Ecology 23:487-505.

Lumsden, N. G., Y. Banerjee, R. M. Kini, S. Kuruppu, and W. C. Hodgson. 2007. Isolation and characterization of rufoxin, a novel protein exhibiting

neurotoxicity from venom of the psammophiine, *Rhamphiophis oxyrhynchus* (Rufous beaked snake). Neuropharmacology 52:1065-1070.

Lynch, V. J. 2007. Inventing an arsenal: Adaptive evolution and neofunctionalization of snake venom phospholipase A_2 genes. BMC Evolutionary Biology 7.

Mackessy, S. P. 1988. Venom ontogeny in the Pacific rattlesnakes *Crotalus viridis helleri* and *C. v. oreganus*. Copeia:92-101.

Mackessy, S. P. 1993a. Fibrinogenolytic proteases from the venoms of juvenile and adult northern Pacific rattlesnakes (*Crotalus viridis oreganus*). Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 106:181-189.

Mackessy, S. P. 1993b. Kallikrein-like and thrombin-like proteases from the venom of juvenile nothern Pacific rattlesnakes (*Crotalus viridis oreganus*). Journal of Natural Toxins 2:223-239.

Mackessy, S. P. 2002. Biochemistry and pharmacology of colubrid snake venoms. Journal of Toxicology-Toxin Reviews 21:43-83.

Mackessy, S. P. 2010a. Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis sensu lato*): Toxicity vs. tenderizers. Toxicon 55:1463-1474.

Mackessy, S. P. 2010b. The field of reptile toxinology: snakes, lizards and their venoms. Pp. 2-23 *in* S. P. Mackessy, ed. Handbook of Venoms and Toxins of Reptiles. CRC Press/Taylor & Francis Group, Boca Raton, FL.

Mackessy, S. P. 2010c. Thrombin-like enzymes in snake venoms. Pp. 519-557 *in* R. M. Kini, M. A. McLane, K. Clemetson, F. S. Markland, and T. Morita, eds. Toxins and Hemostasis: From Bench to Bedside,. Springer-Verlag, Heidelberg.

Mackessy, S. P. and L. Baxter. 2006. Bioweapons synthesis and storage: The venom gland of front-fanged snakes. Zoologischer Anzeiger 245:147-159.

Mackessy, S. P. and W. H. Heyborne. 2010. Cysteine-rich secretory proteins in reptile venoms. Pp. 325-334 *in* S. P. Mackessy, ed. Handbook of Venoms and Toxins of Reptiles. . CRC Press/Taylor & Francis Group, Boca Raton, FL.

Mackessy, S. P., N. M. Sixberry, W. H. Heyborne, and T. Fritts. 2006. Venom of the brown treesnake, *Boiga irregularis*: Ontogenetic shifts and taxa-specific toxicity. Toxicon 47:537-548.

Mackessy, S. P., K. Williams, and A. K. 2003. Characterization of the venom of the midget faded rattlesnake (*Crotalus viridis concolor*): A case of venom paedomorphosis? Copeia 4:769-782.

Macrander, J., M. Broe, and M. Daly. 2015. Multi-copy venom genes hidden in *de novo* transcriptome assemblies, a cautionary tale with the snakelocks sea anemone *Anemonia sulcata* (Pennant, 1977). Toxicon 108:184-188.

Malhotra, A., S. Creer, J. B. Harris, R. Stöcklin, P. Favreau, and R. S. Thorpe. 2013. Predicting function from sequence in a large multifunctional toxin family. Toxicon 72:113-125.

Margres, M., K. Aronow, J. Loyacano, and D. Rokyta. 2013. The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. BMC Genomics 14:1-18.

Margres, M. J., J. J. McGivern, K. P. Wray, M. Seavy, K. Calvin, and D. R. Rokyta. 2014. Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*). Journal of Proteomics 96:145-158.

Martill, D. M., H. Tischlinger, and N. R. Longrich. 2015. A four-legged snake from the early Cretaceous of Gondwana. Science 349:416-419.

Massey, D. J., J. J. Calvete, E. E. Sánchez, L. Sanz, K. Richards, R. Curtis, and K. Boesen. 2012. Venom variability and envenoming severity outcomes of the *Crotalus scutulatus scutulatus* (Mojave rattlesnake) from Southern Arizona. Journal of Proteomics 75:2576-2587.

McGivern, J. J., K. P. Wray, M. J. Margres, M. E. Couch, S. P. Mackessy, and D. R. Rokyta. 2014. RNA-seq and high-definition mass spectrometry reveal the complex and divergent venoms of two rear-fanged colubrid snakes. BMC Genomics 15.

Meier, J. 1986. Individual and age-dependent variations in the venom of the fer-de-lance (*Bothrops atrox*). Toxicon 24:41-46.

Meier, J. and T. A. Freyvogel. 1980. Comparative studies on venoms of the fer-de-lance (*Bothrops atrox)*, carpet viper (*Echis carinatus*) and spitting cobra (*Naja nigricollis*) snakes at different ages. Toxicon 18:661-662.

Menezes, M. C., M. F. Furtado, S. R. Travaglia-Cardoso, A. C. M. Camargo, and S. M. T. Serrano. 2006. Sex-based individual variation of snake venom

proteome among eighteen *Bothrops jararaca* siblings. Toxicon 47:304-312.

Meng, Q.-X., W.-Y. Wang, Q.-M. Lu, Y. Jin, J.-F. Wei, S.-W. Zhu, and Y.-L. Xiong. 2002. A novel short neurotoxin, cobrotoxin c, from monocellate cobra (*Naja kaouthia*) venom: isolation and purification, primary and secondary structure determination, and tertiary structure modeling. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 132:113-121.

Min, X. J., G. Butler, R. Storms, and A. Tsang. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Research 33:W677-W680.

Minton, S. A. 1967. Observations on toxicity and antigenic makeup of venoms from juvenile snakes. Pergamon Press, Oxford.

Minton, S. A. and S. A. Weinstein. 1986. Geographic and ontogenic variation in venom of the western diamondback rattlesnake (*Crotalus atrox*). Toxicon 24:71-80.

Minton, S. A. and S. A. Weinstein. 1987. Colubrid snake venoms: immunologic relationships, electrophoretic patterns. Copeia:993-1000.

Modahl, C. M., R. Doley, and R. M. Kini. 2010. Venom analysis of long-term captive Pakistan cobra (*Naja naja*) populations. Toxicon 55:612-618.

Modahl, C. M., Saviola A.J., and Mackessy S.P. 2015. Venoms of Colubrids. Springer Netherlands.

Montecucco, C., J. M. Gutiérrez, and B. Lomonte. 2008. Cellular pathology induced by snake venom phospholipase A_2 myotoxins and neurotoxins: Common aspects of their mechanisms of action. Cell. Mol. Life Sci. 65:2897-2912.

Mordvintsev, D. Y., Y. L. Polyak, D. I. Rodionov, J. Jakubik, V. Dolezal, E. Karlsson, V. I. Tsetlin, and Y. N. Utkin. 2009. Weak toxin WTX from *Naja kaouthia* cobra venom interacts with both nicotinic and muscarinic acetylcholine receptors. FEBS Journal 276:5065-5075.

Mordvintsev, D. Y., D. I. Rodionov, M. V. Makarova, A. A. Kamensky, N. G. Levitskaya, A. Y. Ogay, D. I. Rzhevsky, A. N. Murashev, V. I. Tsetlin, and Y. N. Utkin. 2007. Behavioural effects in mice and intoxication symptomatology of weak neurotoxin from cobra *Naja kaouthia*. Basic & Clinical Pharmacology & Toxicology 100:273-278.

Mukherjee, A. K. 2007. Correlation between the phospholipids domains of the target cell membrane and the extent of *Naja kaouthia* PLA_2-induced membrane damage: Evidence of distinct catalytic and cytotoxic sites in PLA_2 molecules. Biochimica et Biophysica Acta - General Subjects 1770:187-195.

Mukherjee, A. K. 2010. Non-covalent interaction of phospholipase A_2 (PLA_2) and kaouthiotoxin (KTX) from venom of *Naja kaouthia* exhibits marked synergism to potentiate their cytotoxicity on target cells. Journal of Venom Research 1:37-42.

Mukherjee, A. K. and C. R. Maity. 2002. Biochemical composition, lethality and pathophysiology of venom from two cobras - *Naja naja* and *N. kaouthia*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 131:125-132.

Namiranian, S. and R. C. Hider. 1992. Use of HPLC to demonstrate variation of venom toxin composition in the Thailand cobra venoms *Naja naja kaouthia* and *Naja naja siamensis*. Toxicon 30:47-61.

Nirthanan, S., P. Gopalakrishnakone, M. C. E. Gwee, H. E. Khoo, and R. M. Kini. 2003. Non-conventional toxins from elapid venoms. Toxicon 41:397-407.

Nirthanan, S. and M. C. E. Gwee. 2004. Three-finger alpha-neurotoxins and the nicotinic acetylcholine receptor, forty years on. Journal of Pharmacological Sciences 94:1-17.

Nobuhisa, I., K. Nakashima, M. Deshimaru, T. Ogawa, Y. Shimohigashi, Y. Fukumaki, Y. Sakaki, S. Hattori, H. Kihara, and M. Ohno. 1996. Accelerated evolution of *Trimeresurus okinavensis* venom gland phospholipase A_2 isozyme-encoding genes. Gene 172:267-272.

Ogawa, T., N. Oda, K. Nakashima, H. Sasaki, M. Hattori, Y. Sakaki, H. Kihara, and M. Ohno. 1992. Unusually high conservation of untranslated sequence in cDNA for *Trimeresurus flavoviridis* phospholipase A_2 isozymes. Proceedings of the National Academy of Sciences 89:8557-8561.

Ohkura, K., S. Inoue, K. Ikeda, and K. Hayashi. 1988. Amino-acid sequences of four cytotoxins (cytotoxins I, II, III and IV) purified from the venom of the Thailand cobra, *Naja naja siamensis*. Biochimica et Biophysica Acta 954:148-153.

OmPraba, G., A. Chapeaurouge, R. Doley, K. R. Devi, P. Padmanaban, C. Venkatraman, D. Velmurugan, Q. Lin, and R. M. Kini. 2010. Identification of a novel family of snake venom proteins veficolins from *Cerberus*

rynchops using a venom gland transcriptomics and proteomics approach. Journal of Proteome Research 9:1882-1893.

Ouyang, C. and T.-f. Huang. 1979. α- and β-fibrinogenases from *Trimeresurus gramineus* snake venom. Biochimica et Biophysica Acta - Enzymology 571:270-283.

Pahari, S., S. P. Mackessy, and R. M. Kini. 2007. The venom gland transcriptome of the Desert Massasauga Rattlesnake (*Sistrurus catenatus edwardsii*): Towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). BMC Molecular Biology 8.

Paoletti, A. C., T. J. Parmely, C. Tomomori-Sato, S. Sato, D. Zhu, R. C. Conaway, J. W. Conaway, L. Florens, and M. P. Washburn. 2006. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proceedings of the National Academy of Sciences 103:18928-18933.

Pawlak, J. and R. M. Kini. 2008. Unique gene organization of colubrid three-finger toxins: Complete cDNA and gene sequences of denmotoxin, a bird-specific toxin from colubrid snake *Boiga dendrophila* (Mangrove Catsnake). Biochimie 90:868-877.

Pawlak, J., S. P. Mackessy, B. G. Fry, M. Bhatia, G. Mourier, C. Fruchart-Gaillard, D. Servent, R. Ménez, E. Stura, A. Ménez, and R. M. Kini. 2006. Denmotoxin, a three-finger toxin from the colubrid snake *Boiga dendrophila* (mangrove catsnake) with bird-specific activity. Journal of Biological Chemistry 281:29030-29041.

Pawlak, J., S. P. Mackessy, N. M. Sixberry, E. A. Stura, M. H. Le Du, R. Ménez, C. S. Foo, A. Ménez, S. Nirthanan, and R. M. Kini. 2009. Irditoxin, a novel covalently linked heterodimeric three-finger toxin with high taxon-specific neurotoxicity. FASEB Journal 23:534-545.

Peichoto, M. E., S. P. Mackessy, P. Teibler, F. L. Tavares, P. L. Burckhardt, M. C. Breno, O. Acosta, and M. L. Santoro. 2009. Purification and characterization of a cysteine-rich secretory protein from *Philodryas patagoniensis* snake venom. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 150:79-84.

Peichoto, M. E., F. L. Tavares, G. Dekrey, and S. P. Mackessy. 2011a. A comparative study of the effects of venoms from five rear-fanged snake species on the growth of *Leishmania major*: Identification of a protein with inhibitory activity against the parasite. Toxicon 58:28-34.

Peichoto, M. E., F. L. Tavares, M. L. Santoro, and S. P. Mackessy. 2012. Venom proteomes of South and North American opisthoglyphous (Colubridae and Dipsadidae) snake species: A preliminary approach to understanding their biological roles. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 7:361-369.

Peichoto, M. E., P. Teibler, S. P. Mackessy, L. Leiva, O. Acosta, L. R. Goncalves, A. M. Tanaka-Azevedo, and M. L. Santoro. 2007. Purification and characterization of patagonfibrase, a metalloproteinase showing alpha-fibrinogenolytic and hemorrhagic activities, from *Philodryas patagoniensis* snake venom. Biochim Biophys Acta 1770:810-819.

Peichoto, M. E., B. C. Zychar, F. L. Tavares, L. R. de Camargo Goncalves, O. Acosta, and M. L. Santoro. 2011b. Inflammatory effects of patagonfibrase, a metalloproteinase from *Philodryas patagoniensis* (Patagonia Green Racer; Dipsadidae) venom. Experimental Biology and Medicine 236:1166-1172.

Pierre, L. S., R. Woods, S. Earl, P. P. Masci, and M. F. Lavin. 2005. Identification and analysis of venom gland-specific genes from the coastal taipan (*Oxyuranus scutellatus*) and related species. Cell. Mol. Life Sci. 62:2679-2693.

Pook, C. E. and R. McEwing. 2005. Mitochondrial DNA sequences from dried snake venom: A DNA barcoding approach to the identification of venom samples. Toxicon 46:711-715.

Prado-Franceschi, J. and S. Hyslop. 2002. South American colubrid envenomations. Journal of Toxicology-Toxin Reviews 21:117-158.

Reali, M., F. G. Serafim, M. A. d. Cruz-Höfling, and M. D. Fontana. 2003. Neurotoxic and myotoxic actions of *Naja naja kaouthia* venom on skeletal muscle in vitro. Toxicon 41:657-665.

Reed, L. J. and H. Muench. 1938. A simple method of estimating fifty percent endpoints. American Journal of Tropical Medicine and Hygiene 27:493-497.

Rodrigues, R. S., J. F. da Silva, J. Boldrini Franca, F. P. Fonseca, A. R. Otaviano, F. Henrique Silva, A. Hamaguchi, A. J. Magro, A. S. Braz, J. I. dos Santos, M. I. Homsi-Brandeburgo, M. R. Fontes, A. L. Fuly, A. M. Soares, and V. M. Rodrigues. 2009. Structural and functional properties of Bp-LAAO, a new L-amino acid oxidase isolated from *Bothrops pauloensis* snake venom. Biochimie 91:490-501.

Rodriguez-Robles, J. A. and M. Leal. 1993. *Alsophis portoricensis* (Puerto Rican Racer) diet. Herpetological Review 24:150-151.

Rokyta, D. R., A. R. Lemmon, M. J. Margres, and K. Aronow. 2012a. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13:312-312.

Rokyta, D. R., M. J. Margres, and K. Calvin. 2015. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. G3: Genes, Genomes, and Genetics 5:2375-2382.

Rokyta, D. R., K. P. Wray, A. R. Lemmon, E. M. Lemmon, and S. B. Caudle. 2012b. A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. Toxicon 57:657-671.

Rokyta, D. R., K. P. Wray, and M. J. Margres. 2013. The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. BMC Genomics 14.

Rotenberg, D., E. S. Bamberger, and E. Kochva. 1971. Studies on ribonucleic acid synthesis in the venom glands of *Vipera palaestinae* (Ophidia, Reptilia). Journal of Biochemistry 121:609-612.

Sakurai, Y., H. Takatsuka, A. Yoshioka, T. Matsui, M. Suzuki, K. Titani, and Y. Fujimura. 2001. Inhibition of human platelet aggregation by l-amino acid oxidase purified from *Naja naja kaouthia* venom. Toxicon 39:1827-1833.

Saldarriaga, M. M. a., R. Otero, V. Núñez, M. F. Toro, A. Díaz, and J. M. a. Gutiérrez. 2003. Ontogenetic variability of *Bothrops atrox* and *Bothrops asper* snake venoms from Colombia. Toxicon 42:405-411.

Sanchez, M. N., A. Timoniuk, S. Marunak, P. Teibler, O. Acosta, and M. E. Peichoto. 2014. Biochemical and biological analysis of *Philodryas baroni* (Baron's Green Racer; Dipsadidae) venom: Relevance to the findings of human risk assessment. Human & Experimental Toxicology 33:22-31.

Saviola, A. J., M. E. Peichoto, and S. P. Mackessy. 2014. Rear-fanged snake venoms: An untapped source of novel compounds and potential drug leads. Toxin Reviews 33:1-17.

Schmidt, J. J. and J. L. Middlebrook. 1989. Purification, sequencing and characterization of pseudexin phospholipases A_2 from *Pseudechis porphyriacus* (Australian red-bellied black snake). Toxicon 27:805-818.

Sells, P. G., R. D. G. Theakston, and D. A. Warrell. 1994. Development of α-neurotoxin antibodies in patients envenomed by the monocellate Thai cobra (*Naja kaouthia*). Toxicon 32:1667-1671.

Sunagar, K., T. Jackson, E. Undheim, S. Ali, A. Antunes, and B. G. Fry. 2013. Three-Fingered RAVERs: Rapid Accumulation of Variations in Exposed Residues of Snake Venom Toxins. Toxins 5:2172-2208.

Sunagar, K., E. A. Undheim, H. Scheib, E. C. Gren, C. Cochran, C. E. Person, I. Koludarov, W. Kelln, W. K. Hayes, G. F. King, A. Antunes, and B. G. Fry. 2014. Intraspecific venom variation in the medically significant Southern Pacific Rattlesnake (*Crotalus oreganus helleri*): Biodiscovery, clinical and evolutionary implications. Journal of Proteomics 99:68-83.

Takacs, Z. and S. Nathan. 2014. Animal venoms in medicine. Pp. 252-259 *in* P. Wexler, ed. Encyclopedia of Toxicology (Third Edition). Academic Press, Oxford.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution.

Tan, K. Y., C. H. Tan, S. Y. Fung, and N. H. Tan. 2015. Venomics, lethality and neutralization of *Naja kaouthia* (monocled cobra) venoms from three different geographical regions of Southeast Asia. Journal of Proteomics 120:105-125.

Tan, N.-H., A. Armugam, and P. J. Mirtschin. 1992. The biological properties of venoms from juvenile and adult taipan (*Oxyuranus scutellatus*) snakes. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 103:585-588.

Tan, N.-H., G. Ponnudurai, and P. J. Mirtschin. 1993a. A comparative study of the biological properties of venoms from juvenile and adult inland taipan (*Oxyuranus microlepidotus*) snake venoms. Toxicon 31:363-367.

Tan, N.-H., G. Ponnudurai, and P. J. Mirtschin. 1993b. A comparative study on the biological properties of venoms from juvenile and adult common tiger snake (*Notechis scutatus*) venoms. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 106:651-654.

Tan, N.-H. and C.-S. Tan. 1988. A comparative study of cobra (*Naja*) venom enzymes. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 90:745-750.

Tsetlin, V. 1999. Snake venom alpha-neurotoxins and other 'three-finger' proteins. European Journal of Biochemistry 264:281-286.

Vonk, F. J., N. R. Casewell, C. V. Henkel, A. M. Heimberg, H. J. Jansen, R. J. McCleary, H. M. Kerkkamp, R. A. Vos, I. Guerreiro, J. J. Calvete, W. Wuster, A. E. Woods, J. M. Logan, R. A. Harrison, T. A. Castoe, A. P. de Koning, D. D. Pollock, M. Yandell, D. Calderon, C. Renjifo, R. B. Currier, D. Salgado, D. Pla, L. Sanz, A. S. Hyder, J. M. Ribeiro, J. W. Arntzen, G. E. van den Thillart, M. Boetzer, W. Pirovano, R. P. Dirks, H. P. Spaink, D. Duboule, E. McGlinn, R. M. Kini, and M. K. Richardson. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proceedings of the National Academy of Sciences 110:20651-20656.

Vonk, F. J., K. Jackson, R. Doley, F. Madaras, P. J. Mirtschin, and N. Vidal. 2011. Snake venom: From fieldwork to the clinic. BioEssays 33:269-279.

Wagstaff, S. C., L. Sanz, P. Juarez, R. A. Harrison, and J. J. Calvete. 2009. Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus*. Journal of Proteomics 71:609-623.

Wang, H., X. Chen, L. Wang, W. Chen, M. Zhou, T. Chen, and C. Shaw. 2013. Cloning and characterisation of three novel disintegrin precursors from the venoms of three *Atheris* species: *Atheris chlorechis*, *Atheris nitschei* and *Atheris squamigera*. Toxicon 71:31-40.

Wang, Y. M., J. Parmelee, Y. W. Guo, and I. H. Tsai. 2010. Absence of phospholipase A_2 in most *Crotalus horridus* venom due to translation blockage: comparison with *Crotalus horridus atricaudatus* venom. Toxicon 56:93-100.

Weinstein, S. A., S. A. Minton, and C. E. Wilde. 1985. The distribution among ophidian venoms of a toxin isolated from the venom of the Mojave rattlesnake (*Crotalus scutulatus scutulatus*). Toxicon 23:825-844.

Weinstein, S. A., D. A. Warrell, J. White, and D. E. Keyler. 2011. "Venomous" bites from non-venomous snakes. Elsevier.

Weinstein, S. A., J. White, D. E. Keyler, and D. A. Warrell. 2013. Non-front-fanged colubroid snakes: A current evidence-based analysis of medical significance. Toxicon 69:103-113.

Weisbach, H., A. V. Robertson, B. Witkop, and S. Udenfriend. 1961. Rapid spectrophotometric assays for snake venom l-amino acid oxidase based

on the oxidation of l-kynurenine or 3,4-dehydro-l-proline. Analytical Biochemistry 1:286-290.

Weissbach, H., A. V. Robertson, B. Witkop, and S. Udenfriend. 1960. Rapid spectrophotometric assays for snake venom L-amino acid oxidase based on the oxidation of L-kynurenine or 3,4-dehydro-L-proline. Analytical Biochemistry 1:286-290.

Weldon, C. L. and S. P. Mackessy. 2010. Biological and proteomic analysis of venom from the Puerto Rican Racer (*Alsophis portoricensis*: Dipsadidae). Toxicon 55:558-569.

Weldon, C. L. and S. P. Mackessy. 2012. Alsophinase, a new P-III metalloproteinase with alpha-fibrinogenolytic and hemorrhagic activity from the venom of the rear-fanged Puerto Rican Racer *Alsophis portoricensis* (Serpentes: Dipsadidae). Biochimie 94:1189-1198.

Wijeyewickrema, L. C., E. E. Gardiner, Y. Shen, M. C. Berndt, and R. K. Andrews. 2007. Fractionation of snake venom metalloproteinases by metal ion affinity: A purified cobra metalloproteinase, Nk, from *Naja kaouthia* binds Ni^{2+}-agarose. Toxicon 50:1064-1072.

Wooldridge, B. J., G. Pineda, J. J. Banuelas-Ornelas, R. K. Dagda, S. E. Gasanov, E. D. Rael, and C. S. Lieb. 2001. Mojave rattlesnakes (*Crotalus scutulatus scutulatus*) lacking the acidic subunit DNA sequence lack Mojave toxin in their venom. Comparative Biochemistry and Physiology - Part B: Biochemistry and Molecular Biology 130:169-179.

Wüster, W. 1996. Taxonomic changes and toxinology: Systematic revisions of the asiatic cobras (*Naja naja* species complex). Toxicon 34:399-406.

Yamazaki, Y. and T. Morita. 2004. Structure and function of snake venom cysteine-rich secretory proteins. Toxicon 44:227-231.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586-1591.

Yates, C. M., I. Filippis, L. A. Kelley, and M. J. E. Sternberg. 2014. SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. Journal of Molecular Biology 426:2692-2701.

Yee, J. S. P., N. L. Gong, F. Afifiyan, D. H. Ma, P. S. Lay, A. Armugam, and K. Jeyaseelan. 2004. Snake postsynaptic neurotoxins: Gene structure, phylogeny and applications in research and therapy. Biochimie 86:137-149.

Zelanis, A., A. K. Tashima, M. M. T. Rocha, M. F. Furtado, A. C. M. Camargo, P. L. Ho, and S. M. T. Serrano. 2009. Analysis of the ontogenetic variation in the venom proteome/peptidome of *Bothrops jararaca* reveals different strategies to deal with prey. Journal of Proteome Research 9:2278-2291.

Zelanis, A., M. M. Teixeira da Rocha, and M. d. F. Domingues Furtado. 2010. Preliminary biochemical characterization of the venoms of five Colubridae species from Brazil. Toxicon 55:666-669.

Zhang, B., Q. H. Liu, W. Yin, X. W. Zhang, Y. J. Huang, Y. F. Luo, P. X. Qiu, X. W. Su, J. Yu, S. N. Hu, and G. M. Yan. 2006. Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. BMC Genomics 7.

Zhang, J. 2003. Evolution by gene duplication: An update. Trends in Ecology & Evolution 18:292-298.

Zhang, J., K. Kobert, F. T., and A. Stamatakis. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics (Oxford, England) 30:614-620.

Zhang, Z., X. Zhang, T. Hu, W. Zhou, Q. Cui, J. Tian, Y. Zheng, and Q. Fan. 2015. Discovery of toxin-encoding genes from the false viper *Macropisthodon rudis*, a rear-fanged snake, by transcriptome analysis of venom gland. Toxicon 106:72-78.

Zybailov, B., A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens, and M. P. Washburn. 2006. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. Journal of Proteome Research 5:2339-2347.

APPENDIX A

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE (IACUC) APPROVAL

IACUC Memorandum

To: Dr. Steve Mackessy

From: Laura Martin, Director of Compliance and Operations

CC: IACUC Files

Date: 2/11/2015

Re: IACUC Protocol **1504D-SM-SMLBirds-18** Approval

The University of Northern Colorado Institutional Animal Care and Use Committee has completed a final review of your protocol "**Toxicity of Venoms and Purified Toxins to Mice, Birds & Lizards**". The protocol review was based on the requirements of Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training; the Public Health Policy on Humane Care and Use of Laboratory Animals; and the USDA Animal Welfare Act and Regulations. Based on the review, the IACUC has determined that all review criteria have been adequately addressed. The PI/PD is approved to perform the experiments or procedures as described in the identified protocol as submitted to the Committee. This protocol has been assigned the following number **1504D-SM-SMLBirds-18**.

The next annual review will need to be approved prior to February 11, 2016.

Sincerely,

Laura Martin, Director of Compliance and Operations

IACUC Memorandum

To: Steve Mackessy

From: Laura Martin, Director of Compliance and Operations

CC: IACUC Files

Date: 4/28/2015

Re: IACUC Protocol Approval 1302D-SM-S-16

The UNC IACUC has completed a final review of your protocol "Analysis of Venoms from Viperid Snakes- Biochemical Composition, Activities".

The committee's review was based on the requirements of the Government Principles, the Public Health Policy, the USDA Animal Welfare Act and Regulations, and the Guide for the Care and Use of Laboratory Animals, as well as university policies and procedures related to the care and use of live vertebrate animals at the University of Northern Colorado.

Based on the review, the IACUC has determined that all review criteria have been adequately addressed. The PI/PD is approved to perform the experiments or procedures as described in the identified protocol as submitted to the Committee. **This protocol has been assigned the following number 1302D-SM-S-16.**

The next annual review will be due before **May 25, 2015**.